REC'D 12 AUG 1999
WIPO PCT

IA 123343

THE UNITED STATES OF AMERICA

TO ALL TO WHOM THESE PRESENTS SHALL COME:

UNITED STATES DEPARTMENT OF COMMERCE

United States Patent and Trademark Office

July 14, 1999 0k99/408

THIS IS TO CERTIFY THAT ANNEXED HERETO IS A TRUE COPY FROM THE RECORDS OF THE UNITED STATES PATENT AND TRADEMARK OFFICE OF THOSE PAPERS OF THE BELOW IDENTIFIED PATENT APPLICATION THAT MET THE REQUIREMENTS TO BE GRANTED A FILING DATE UNDER 35 USC 111.

APPLICATION NUMBER: ***60/094,868***

FILING DATE: ***July 29, 1998***

PRIORITY DOCUMENT

SUBMITTED OR TRANSMITTED IN COMPLIANCE WITH RULE 17.1(a) OR (b)

By Authority of the
COMMISSIONER OF PATENTS AND TRADEMARKS

H. Phillips

H. PHILLIPS
Certifying Officer

Box Provisional Application
PTO/SB/16(6-95)

PROVISIONAL APPLICATION FOR PATENT COVER SHEET

This is a request for filing a PROVISIONAL APPLICATION FOR PATENT under 37 C.F.R. § 1.53 (c).

Filing Date	July 29, 1998	Docket Number	0459-0252P

INVENTOR(s)/APPLICANT(s)

LAST NAME	FIRST NAME	MIDDLE INITIAL	RESIDENCE (CITY AND EITHER STATE OR FOREIGN COUNTRY)
HALKIER	Torben		Birkerød, Denmark
JESPERSEN	Lene		Copenhagen V, Denmark
JENSEN	Allan		Copenhagen S, Denmark

TITLE OF THE INVENTION (280 characters max)

NOVEL METHODS FOR THE IDENTIFICATION OF LIGAND AND TARGET BIOMOLECULES

CORRESPONDENCE ADDRESS

Birch, Stewart, Kolasch & Birch, LLP
P.O. Box 747
Falls Church

STATE	VA	ZIP CODE	22040-0747	COUNTRY	U.S.A.

ENCLOSED APPLICATION PARTS *(check all that apply)*

X Specification *Number of Pages* 86 ___ Small Entity Statement
X Drawing(s) *Number of Sheets* 1 ___ Other (specify) ___

METHOD OF PAYMENT *(check one)*	PROVISIONAL FILING FEE
X A check or money order is enclosed to cover the Provisional filing fees	___ Small Entity ($75.00)
X The Commissioner is hereby authorized to charge filing fees and credit Deposit Account Number 02-2448, if necessary.	X Large Entity ($150.00)

The invention was made by an agency of the United States Government or under a contract with an agency of the United States Government.

X No.

___ Yes, the name of the U.S. Government agency and the Government contract number are: ___

Respectfully submitted,

BIRCH, STEWART, KOLASCH & BIRCH, LLP

By [signature] #25977 Date July 29, 1998

Leonard R. Svensson
Reg. No. 30,330
P.O. Box 747
Falls Church, VA 22040-0747
(703) 205-8000

___ Additional inventors are being named on separately numbered sheets attached hereto.

Rev. 1/20/98

NOVEL METHODS FOR THE IDENTIFICATION OF LIGAND AND TARGET BIOMOLECULES

FIELD OF THE INVENTION

The present invention pertains to a novel method for the identification/preparation of peptides or ribonucleic acids capable of modulating the activity *in vivo* of target enzymes in eukaryotic cells. More specifically, the invention provides a method for identification/preparation of hitherto unknown enhancers as well as inhibitors of *in vivo* enzyme activity in eukaryotic cells. Furthermore, the invention relates to methods for identification of unknown interactions (*i.e.* identification of a target and/or a ligand but also of hitherto unknown interactions between known ligands and known targets). These novel methods employ enzyme inhibitor structures as scaffolds in order to intracellularly present potentially biologically active peptides or ribonucleic acids in a stable form. Also disclosed herein are methods for the preparation of the hitherto unknown ligands or targets as well as methods for the preparation of vectors and transformed cells carrying the genetic information encoding these ligands and targets. Finally, the invention relates to a method for the preparation of a medicinal product which is based on initial identification of targets or ligands according to the present invention.

BACKGROUND OF THE INVENTION

The Cell Screen technique is a method that allows for the identification of peptide sequences having biological activity *in vivo* and is disclosed in WO 96/38553. In short, libraries of random peptides are expressed intracellularly in mammalian cells, such that one cell expresses one single or a few heterologous short peptides. Cells that change a preselected phenotype under certain conditions can be isolated and the peptide that they express can hence be identified. The

intracellular component with which the peptide interacts (the target molecule) may subsequently be obtained using e.g. affinity columns carrying the immobilized synthetic peptide.

Although the CellScreen technology has shown great promise for identifying new drug targets, it is an inherent problem that the intracellular environment is relatively hostile to many expression products which are expressed. In other words, interesting peptide or nucleic acid sequences which potentially are capable of interacting with an important target molecule may be degraded or inactivated inside the cell before any effect on phenotype can be detected.

OBJECT OF THE INVENTION

It is an object of the invention to provide improvements in the CellScreen technology by overcoming the above-mentioned problems of potential instability of expressed sequences. Furthermore, it is an object of the invention to expand the utility of the CellScreen technology to also encompass screening in prokaryotic cells.

SUMMARY OF THE INVENTION

A significant number of enzyme activity modulators of plant, microbial and eukaryotic cell origin have been described, cf. below.

Since many of the naturally occurring processes inside cells are regulated by enzymes, the inventors disclose herein a method for expression of large intracellular libraries of such enzyme activity modulators in which the active site of said enzyme activity modulators have been altered by introduction of stretches of randomized amino acid sequences or by introduction of random nucleotides at specific sites in the active site. This

creates libraries of putative modulators capable of modulating the activity of an array of different enzymes inside cells. By expressing these modulators in cell lines according to a novel variation of the CellScreen technology the enzymatic regulatory mechanisms inside the individual cell in said cell line will be affected differently leading to different phenotypic properties such as e.g. resistance towards hypoglycemia, cytokine killing, toxic compounds, virus infection etc.

The main advantage of using known enzyme activity modulators such as enzyme *inhibitors* as scaffolds is that many of these in their native form are stable in the intracellular environment. The problem of using e.g. antibody fragments as scaffolds for intracellular presentation of random peptides is that many such antibody fragments are susceptible to the proteolytic and reducing intracellular environment and therefore such molecules can, if carefully tailored, maintain their intracellular stability and at the same time incorporate random sequences which are screened for biological activity. Furthermore, the effectivity of such a screen for biologically active substances will be higher than if using an unstable scaffold (such as a e.g. a coiled coil structure) or no scaffold at all, since none or only a very limited number of the randomized sequences will be degraded before they can exert their effects in the cells. Finally, a majority of such enzyme activity modulators have an active site which is the perfect position in the molecule to modify, since the active site is normally presented in a stable configuration to the environment.

Hence, one aspect of the invention pertains to a method for identifying an *in vivo* active modulator of activity of a target enzyme, the method comprising the steps of (a) preparing a pool of expression vectors, each vector of said pool containing at least one member from a library of randomly modified nucleotide sequences derived from a parent nucleotide sequence

encoding a parent peptide or parent ribonucleic acid which modulates the target enzyme activity, (b) transforming a population of substantially identical cells with said vectors of said pool so as to obtain transformed cells, said substantially identical cells being ones which harbour the target enzyme, (c) culturing said transformed cells under conditions facilitating expression of said randomly modified nucleotide sequences, (d) examining said transformed cells and isolating transformed cell(s) wherein the activity of the target enzyme is modulated, and (e) determining said randomly modified nucleotide sequence of said vector present in cell(s) isolated in step (d) and/or determining the amino acid sequence or the ribonucleic acid sequence of the expression product encoded by said randomly modified nucleotide sequence.

In this part of the invention it is normally preferred that the randomly modified nucleotide sequences consist of 1) an invariable part of the parent nucleotide sequence, and 2) random nucleotides. In line with the above, the invariable part of the parent nucleotide sequence preferably encodes a scaffold portion of the parent peptide or of the parent ribonucleic acid which serves to stabilize said polypeptide fragment or ribonucleic acid fragment.

As mentioned above, the use of enzyme modulator scaffolds also provides for expression of *stable*, biologically active modulators which interact with other biomolecules than enzymes. In other words, in cases where random nucleotides are inserted in a scaffold structure, the outcome will be a stable expression product, the activity of which does not necessarily have anything to do with enzyme activity modulation.

Hence, another part of the invention is a method for identifying a modulator in the form of a biologically active polypeptide fragment or ribonucleic acid fragment which is capable of de-

tectably modulating, *in vivo*, a phenotypic trait in a cell, the method comprising the steps of

(a) preparing a pool of expression vectors, each vector of said pool containing at least one member from a library of randomly modified nucleotide sequences derived from a parent nucleotide sequence encoding a parent peptide or parent ribonucleic acid which *in vivo* modulates activity of a known enzyme, wherein the randomly modified nucleotide sequences comprise
an invariable part encoding a scaffold portion of the parent peptide or of the parent ribonucleic acid, said scaffold portion serving to stabilize said polypeptide fragment or ribonucleic acid fragment, and
- random nucleotides,

(b) transforming a population of substantially identical cells with said vectors of said pool so as to obtain transformed cells,

(c) culturing said transformed cells under conditions facilitating expression of said randomly modified nucleotide sequences,

(d) examining said transformed cells and isolating transformed cell(s) wherein the preselected phenotypic trait is modulated by the presence of the expressed randomly modified nucleotide sequence, and

(e) determining said randomly modified nucleotide sequence of said vector present in cell(s) isolated in step (d) and/or determining the amino acid sequence or the ribonucleic acid sequence of the expression product encoded by said randomly modified nucleotide sequence.

Finally, the invention also pertains to the general use of intracellularly stable scaffold proteins, ribonucleotides, or

fragments thereof for the presentation of random sequences in the CellScreen technology. As mentioned above, although the concept of using scaffold molecules has been discussed in the prior art, the issue of the *stability* of the scaffold system has not been detailed.

The stability and usefulness of a putative intracellular scaffold is dependent on a number of factors. First of all, it is essential that the relevant cell wherein the scaffold is to be expressed is capable of expressing the scaffold molecule in a functional form; that is, in prokaryotic systems some eukaryotic proteins will not fold correctly, hence rendering the use of such a protein unsuitable as a scaffold in that type of cell. Second, the scaffold should be relatively resistant to the reducing and catalytic environment inside intact cells. However, even when a scaffold molecule is relatively susceptible to the inactivating nature of the intracellular environment, this can be remedied if the production rate of the scaffold molecule is sufficiently high.

In steady state, the intracellular concentration of a scaffold molecule will be a function of the following formula:

$$C_{scaffold} = \frac{R_p}{R_d}$$

-where R_p is the rate of production of the scaffold molecule (moles $\cdot$ s^{-1}) and R_d is the inactivation constant for the scaffold molecule (l $\cdot$ s^{-1}), *i.e.* the rate of inactivation of the scaffold molecule is determined by $\frac{\delta M}{\delta t} = R_d C_{scaffold}$ (which, in steady state, of course equals R_p. In other words, when assessing the suitability of a potential scaffold molecule, it should according to the present invention be tested whether the molecule can be kept at a sufficiently high concentration inside the relevant cell wherein the CellScreen test is going to be carried out.

Therefore, a very broad aspect of the invention pertains to a method for identifying a modulator in the form of a biologically active polypeptide fragment or ribonucleic acid fragment which is capable of detectably modulating, in vivo, a phenotypic trait of a cell, the method comprising the steps of

(a) preparing a pool of expression vectors, each vector of said pool containing at least one member from a library of randomly modified nucleotide sequences derived from a parent nucleotide sequence encoding a parent peptide or parent ribonucleic acid which is stable intracellularly, wherein the randomly modified nucleotide sequences comprise

- an invariable part encoding a scaffold portion of the parent peptide or of the parent ribonucleic acid, said scaffold portion serving to stabilize said polypeptide fragment or ribonucleic acid fragment, and
- random nucleotides,

(b) transforming a population of substantially identical cells with said vectors of said pool so as to obtain transformed cells,

(c) culturing said transformed cells under conditions facilitating expression of said randomly modified nucleotide sequences,

(d) examining said transformed cells and isolating transformed cell(s) wherein the preselected phenotypic trait is modulated by the presence of the expressed randomly modified nucleotide sequence, and

(e) determining said randomly modified nucleotide sequence of said vector present in cell(s) isolated in step (d) and/or determining the amino acid sequence or the ribonucleic acid sequence of the expression product encoded by said randomly modified nucleotide sequence,

wherein the expression product of the nucleic acid sequence which encodes the intracellularly stable parent peptide or parent ribonucleic acid is one which, when produced by the substantially identical cells, is present in an effective concentration and in a functional state.

In other words, it is essential that the suitability of the scaffold molecule is evaluated prior to performing the steps of the CellScreen technology in order to confirm that the scaffold molecule in unmodified form can be expressed and maintained at a sufficiently high concentration/activity in the cellular system where the method of the invention is to be exercised.

According to WO 96/38553, the isolation of a drug target molecule can be made more efficient if the random peptide sequences are inserted into larger polypeptides functioning as scaffolds for display of the random amino acid sequences. Such scaffolds would probably also lead to higher affinity interaction with the target molecule.

Nothing is, however, mentioned about the use of scaffolds derived from naturally occurring protein inhibitors of enzymes. Inhibition of enzymatic activity by such inhibitors - as opposed to the simple binding of a target protein inside a cell as was suggested in the CellScreen technology - is a much more efficient way to affect intracellular biochemical events.

DETAILED DISCLOSURE OF THE INVENTION

Definitions

In the following, a number of terms will be defined for the purposes of the present disclosure:

A "modulator" is in the present context a biomolecule which, when expressed *in vivo*, to effect the activity of another biomolecule in the cell. Thus, the modulator in essence can inhibit or enhance the activity of the biomolecule. Furthermore, the modulator can interact directly with the biomolecule, but the effect might as well be indirect (also known as transdominant), i.e. the activity change of the biomolecule is brought about by changes in the cell's biochemical machinery, changes which are ultimately the result of the presence of the expression product of the randomized nucleic acid sequence.

A "randomly modified nucleotide sequence" is a nucleotide sequence which in a number of positions has been subjected to insertion or substitution by nucleotides, the nature of which cannot be predicted. In many cases the random nucleotides or nucleotide sequences inserted will be "completely random" (e.g. as a consequence of randomized synthesis or PCR-mediated mutagenesis). However, as will appear from the disclosure below, the random sequences can also include sequences which have a common functional feature (e.g. reactivity with a ligand of the expression product) or the random sequences can be random in the sense that the ultimate expression product is of completely random sequence with e.g. an even distribution of the different amino acids.

"Substantially identical cells" is a term herein intended to designate cells which all exhibit a specific phenotypic trait in such a manner that a change in the expression of said trait in one cell due to an interaction effected by the introduction of random nucleotides according to the inventive methods would also occur in one of the other substantially identical cells had these been transformed with the same vector. In other words, the important parameter to assess when choosing substantially identical cells in the inventive methods is whether an observed change in one cell's exhibition of the phenotypic trait can be

taken as an indication that any other cell in the population would have behaved the same way as a consequence of the same change. Hence, substantially identical cells can be clonal cells or cells of a cell line or they can be cells of a cell culture or a tissue culture.

A "phenotypic trait" is the observable result of a certain gene composition in a cell (genotype), *i.e.* a property of a cell (detected by chemical, physical, immunological or any other suitable means) which depends on the presence of one or several genes and the expression rate thereof. Thus, the phenotypic trait can be any of a number of different properties: activity of an enzyme, effects of interaction between receptors and ligands, cell survival rate, presence or absence of an antigen, expression rate, etc.

"Peptide" is in the present context intended to mean both short peptides of from 2 to 10 amino acid residues, oligopeptides of from 11 to 100 amino acid residues, and polypeptides of more than 100 amino acid residues. Furthermore, the term is also intended to include proteins, *i.e.* functional biomolecules comprising at least one polypeptide; when comprising at least two polypeptides, these may form complexes, be covalently linked, or may be non-covalently linked. The polypeptide(s) in a protein can be glycosylated and/or lipidated and/or comprise prosthetic groups.

"*In vivo*" is herein mean to designate the environmental conditions inside living cells (*i.e.* cells which are metabolically active and can maintain their vital functions); the living cells may be kept in culture or may be present in a natural habitat (e.g. functioning as part of a larger, multicellular organism). Thus, the term "*in vivo*" also refers to *in vitro* culturing of cells as long as the effect being observed is taking place in the living cell.

"Transformation": A process by which the genetic material carried by an individual cell is altered by incorporation of exogenous DNA into its genome.

"Transfection": The uptake, incorporation, and expression of recombinant DNA by cells.

"Transduction": The transfer of genetic information from one cell to another by way of a viral vector.

The term "effective part" when used in the context of a protein, peptide, or ribonucleic acid is in the present context intended to mean a part (e.g. a subsequence or, in the case of an n-meric protein, a less-than-n-meric molecule) which has retained the desired functionality of the native molecule from which the protein or peptide is derived. For instance, in the case of CI-2A only the truncated form of the molecule described in Example 1 seems to be necessary to ensure expression of the active inhibitor intracellularly, and hence the truncated form of that molecule constitutes an effective part of CI-2A.

General properties of the inventive method

In general, the disclosures in WO 96/38553 and WO 97/27212 relating

to preparation of randomized sequences, to the choice of fusion partners (except for the choice of scaffolds) for the random sequences, to the choice and composition of targeting sequences, to the choice of nucleic acids to be randomly modified, to the choice of randomization method, to the methods of introducing the nucleic acids into the relevant cell type, to the choice of (retroviral) vectors (where applicable), to the method of producing the vectors, the choice of promoters, to the choice of packaging cells

(where applicable), to the methods of concentrating infectious virions from the packaging cells (where applicable), to the choice of substantially identical cells to use in the method, to the type of phenotypic changes detected, to the manner in which the change is detected, to methods of isolating the phenotypically changed cells, to the isolation and sequence determination of the randomly modified sequences, to the isolation and characterization of the target for the randomized product, to screening methods, and to the choice of applications of the methods

also are relevant for the purposes of the present invention. Therefore, the disclosures of these two patent applications are hereby incorporated by reference herein. However, since both of these references are focussed on the use of the general principle in higher eukaryotic cells, the present disclosure will also detail on embodiments pertaining to the use in prokaryotic systems, cf. below. However, the more general part of the two above-referenced disclosures which without any difficulty for the skilled person could be applied in the context of a prokaryotic system are also regarded as relevant and important embodiments of the present invention insofar as it relates to the use of the methods in prokaryotic systems.

It is normally preferred that the transformed cells which are being examined in step (d) predominantly carries (and expresses) one single copy of the vector. By ensuring this, the interpretation of a change in phenotype of the cells becomes a much easier task, whereas the interpretation of a phenotype change in cells expressing more than one single randomized sequence renders unclear which of the transforming vectors is responsible for the change.

To ensure that predominantly one vector has transformed each of the cells examined it is e.g. feasible that the transformation

step (b) is performed under such conditions that the cells transformed are predominantly or at most transformed with one single vector from said pool (this can e.g. be achieved by adjusting the concentration of infectious virions in embodiments of the invention where the transformation is obtained by means of transduction), or wherein, prior to carrying out step (d), cells being transformed with more than one vector from said pool are substantially excluded from the further steps. This latter option requires that it is possible to quantify the number of transforming vectors and this can be achieved by including a detectable marker in the expression product, e.g. a flourescent probe. Another option is to rescreen cells which exhibit changes in phenotype, thereby ascertaining whether more than one vector has transformed the cell.

It will be understood that the molecule chosen for the purpose of being a scaffold, and wherein the random sequences are ultimately introduced, can be either a peptide sequence or a nucleic acid sequence, such as an RNA fragment interfering (in an antisense-manner) with mRNA, tRNA or with a ribozyme. Alternatively, such an RNA fragment could exhibit ribozyme activity itself, thereby having an indirect influence on the expression rate of other enzymes. At any rate, the resulting product, i.e. the randomized expression product, can be a peptide or a ribonucleic acid such as a ribozyme.

One important feature of the scaffold is, as mentioned above, that it is stable towards proteolytic attack and/or is insensitive to a reducing environment, such as the one which is found intracellularly.

In preferred embodiments of the invention the random nucleotides are introduced in part(s) of the parent nucleotide sequence which encode(s) the active site(s) of the parent peptide or parent ribonucleic acid, or the part(s) which encode(s) struc-

ture(s) interfering with the active site(s). As discussed above, the active site (as well as other exposed structures of the scaffold) need to be stable presented to the environment in order to be able to interact with other biomolecules. Hence, preferably the *invariable* part of the nucleotide sequence encodes truncated parts of the parent peptide or parent ribonucleic acid sufficient to maintain stability of the randomized product.

In some embodiments it is preferred that the invariable part of the parent nucleotide sequence encodes a peptide which is free from disulphide bridges. This is, as mentioned in Example 2 below, due to the fact that disulphide bridges are not formed in the nucleus or in the cytosol. Hence, in cases were the scaffold must be in a functional state when present in the nucleus or the cytosol, it would normally be preferable to use a scaffold which does not contain disulphide bridges or which do not rely on these in order to maintain stability and functionality. On the other hand, in embodiments where it is desired that the randomized expression product is confined to the ER, or to another compartment allowing for the presence of stable disulphide bridges, before the randomized sequence is presented to the environment in a satisfactory manner, it is of course desirable that the invariable part of the parent nucleotide sequence encodes a peptide having disulphide bridges, because the chances of having a correctly folded and functional scaffold outside such compartments is relatively small.

It will be understood that the random nucleotides are preferably introduced in the form of an insertion or a substitution into the parent nucleotide sequence, optionally in combination with deletion(s) in the parent nucleotide sequence. Deleted sequences in the parent polypeptide could e.g. be parts of an active site, the presence of which in unaltered form is toxic or otherwise deleterious to the transformed cells.

The number of random nucleotides introduced can vary to a great degree but normally the number is between 3 and about 100. In this range it is preferred that at least 5, such as at least 7, and better, at least 9-12 random nucleotides are introduces. On the other hand, it is preferred that at most 90, such as at most 70 or 80 random nucleotides are introduced. The most preferred number of introduced randomized nucleotides varies between 15-60, preferably 20-55 or 25-50 nucleotides. Especially preferred numbers of random nucleotides are 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, and 60 nucleotides.

The random nucleotides are introduced in the scaffold in the form of nucleotide sequences and/or in the form of single random nucleotides introduced at specific sites in the parent nucleotide sequence. A variation is to substitute a part of the scaffold sequence with a sequence which retains parts of the scaffold sequence (e.g. those which are essential for stability/functionality) but where other parts are randomized.

The random nucleotides are preferably selected from the group consisting of

- synthetic, completely random deoxyribonucleotides;
- synthetic random DNA sequences, wherein limitation on randomization of some nucleotides is introduced so as to limit the number of available sequences and/or to avoid undesired stop codons and/or to facilitate introduction of post-translational modifications of expressed peptide(s);
- synthetic random DNA sequences as in (1) or (2) coupled to a sequence encoding a purification tag; and
- CDR encoding nucleotide sequences isolated from a library of immune-competent cells raised against an antigen (in this embodiment it is preferred that CDR encoding nucleotide sequences encode CDR-3 peptide sequences).

The latter type of "randomization" actually introduces a restriction on randomness which ensures that the sequences introduced encodes an antigen recognizing region. It is, however, well known that a polyclonal immune response against an antigen consist of a large number of immune competent cells which all react with the same antigen (or perhaps even with the same epitope) but the correlation between amino acid sequences of the CDRs and recognition of the epitope(s) is virtually impossible to deduce.

An alternative way to introduce limitations on the randomness of the nucleic acid sequences which are ultimately tested in the substantially identical cells is the following: Upon preparation of the vectors, they are used in a 1st round of phage display, where the phages transformed with the vectors are panned against a library containing a ligand of choice. As for the technique of employing CDR encoding sequences, the result is that the sequences which are ultimately tested in the substantially identical cells are "unpredictable" (and thereby random) but nevertheless selected on the basis of a functional feature. Again, the lack of known correlation between nucleic acid sequences and the interaction in three-dimensional space between the expression product and a ligand of choice has the consequense that the tested subgroup of sequences still is randomized.

In a special embodiment of the above-technique where the method of the invention is combined with phage display, both test systems are repeated in an alternating manner, that is a shuffling between intracellular expression in the substantial identical cells and panning of a phage library.

In order to obtain an controlled distribution of amino acids in the randomized peptides, when the modulator is a peptide, it is practical that the random nucleotides are prepared by codon

split synthesis where defined DNA codons are synthesized in a random order; a thorough description of this principle is given in WO 96/33553. The preferred embodiment in this context is one wherein the relative amount of synthesized codons ensure that all encoded amino acids will be present with substantially the same frequency in encoded polypeptide fragments.

In order to introduce the randomized fragments properly into the vectors, it is according to the invention preferred that the random nucleotides are introduced into the expression vector by the principle of site directed PCR-mediated mutagenesis. However, other options are known to the skilled person, and it is e.g. possible to insert synthetic random sequence libraries into the vectors as well.

Apart from having the randomized fragment of the expression product introduced into a scaffold in accordance with the present invention, it is often necessary to couple the random sequence to a fusion partner by having the randomized nucleotide sequence fused to a nucleotide sequence encoding at least one fusion partner. Such a fusion partner can e.g. facilitate expression and/or purification/isolation and/or further stabilization of the expression product.

For the purposes of purification, the fusion partner can include a purification tag such as His_6 tag, myc tag, BSP biotinylation target sequence, of BirA, flu tag, lacZ, and GST. Furthermore, the fusion partner may include a sorting signal or a targeting sequence, cf. the discussions below.

In embodiments where the modulator is itself a modulator of enzyme activity, it is in theory possible to effect both the K_M and/or the V_{max} of the relevant enzyme. A reduction in K_M results in less effectivity of the relevant enzyme insofar that an increased substrate concentration is required to obtain 50% of

maximum activity of the enzyme. An increase of K_M has the opposite effect. Of course, interference with an enzyme which effects V_{max} has as a consequence that the maximum possible rate of activity of the enzyme is increased (when V_{max} is increased) or decreased (when V_{max} is increased). At any rate, the modulator of the enzyme activity will give the phenotypic impression that the enzyme activity has either been inhibited or stimulated. It is preferred in this embodiment that the modulator is an inhibitor.

When the method of the invention has finally lead to the identification of a modulator or of a target therefor, it is preferred that the 3-dimensional structure of the identified modulator is resolved, since this allows for the implementation of rational drug screening and computer drug modelling methods.

Use of the inventive methods in prokaryotic systems

The originally envisaged technology disclosed in WO 96/38553 focussed on screening for interactions in *eukaryotic* cells. However, the technology is also applicable in prokaryotic systems.

For instance, it is expected that the present invention will allow for identification of hitherto unknown interactions in pathogenic bacteria, interactions which will be useful in the course of developing new antibiotics. Since the inventive methods allows for the identification of both novel ligand peptides and ribonucleic acids as well as of the target molecules for these ligands, the investigator is provided with the necessary tools for instigating computer drug modelling and for performing traditional drug screening, once such ligands and/or targets have been identified/isolated.

However, apart from the approach of identifying antibacterial effects and substances, the method also opens up for improve-

ments in industrial fermentation processes. In such cases it will e.g. be possible to identify biomolecules which are important in the biochemical pathways in lactic acid bacteria and thereby provide tools for the production of new dairy products such as cheese, yoghourt, and other products of lactic acid bacterial fermentation.

Somewhat related to this approach is the use of the methods of the present invention in screening performed on bacterial cultures used in purification processes. It is well-known in the art of e.g. waste water purification that the microbiological cultures (activated sludge) which conduct the degradation of organic material, are relatively vulnerable *vis à vis* changes in the environment and therefore the provision of more robust strains of bacteria would be one way to improve such systems. Alternatively, the method of the present invention would also allow for the identification/isolation of ligands and targets in such bacteria which, when interacted with, can lead to e.g. increased efficacy in degradation of specific organic or inorganic substrates.

As will be appreciated from the above, the present invention therefore is highly useful in prokaryotic systems.

For the purposes of using the method of the invention in prokaryotic cells, it is preferred that the prokaryotic cells are bacteria selected from the group consisting of *Bacillus* spp. (e.g. *B. anthracis*, *B. subtilis* and *B. cereus*), *Clostridium* spp. (e.g. *C. botulinum*, *C. difficile*, *C. perfringens*, and *C. tetani*), *Corynebacterium* spp. (e.g. *C. diphtheriae*, and *C. pyogenes*), *Staphylococcus* spp. (e.g. *S. aureus* and *S. albicans*), *Streptococcus* spp. (e.g. *S. pneumoniae*, *S. pyogenes*, and *S. agalactiae*), *Escherichia coli*, *Serratia marcescens*, *Klebsiella* spp. (e.g. *K. pneumoniae*), *Proteus* spp. (e.g. *P. mirabilis*), *Citrobacter* spp. (e.g. *Citrobacter freundii*), *Salmonella* spp.

(e.g. *S. typhi*, *S. typhimurium*, *S. shottmuelleri* and *S. paratyphi*), *Shigella* spp. (e.g. *S. dystenteriae*, *S. flexneri*, *S. boydii*, and *S. sonnei*), *Pseudomonas* spp. (e.g. *P. aeruginosa*, *P. pseudomallei*, and *P. mallei*), *Acinetobacter* spp., *Aeromonas* spp., *Plesiomonas* spp., *Yersinia* spp. (e.g. *Y. pestis*, *Y. enterocolitica*, and *Y. pseudotuberculosis*), *Francisella tularensis*, *Vibrio* spp. (e.g. *V. cholerae* and *V. parahaemolyticus*), *Campylobacter* spp. (e.g. *C. jejuni* and *C. coli*), *Helicobacter pylori*, *Haemophilus* spp. (e.g. *H. influenzea*, *H. parainfluenzae*, and *H. aegyptius*), *Bordetella* spp. (e.g. *B. pertussis*, *B. parapertussis*, and *B. bronchiseptica*), *Brucella* spp., *Neisseria* spp. (e.g. *N. gonnorhoeae* and *N. meningitidis*), *Treponema pallidum*, *Leptospira interrogans*; *Borrelia* spp. (e.g. *B. burgdorferi sensu stricto*, *B. garinii*, *B. afzelii*, and *B. recurrentis*), *Legionella pneumophila*, *Listeria monocytogenes*, *Mycobacterium* spp. (e.g. *M. tuberculosis*, *M. bovis*, *M. africanum*, *M. kansasii*, and *M. leprae*), *Treponema pallidum*, *Chlamydia trachomatis*, *Actinomyces* spp., *Rickettsia* spp., and *Mycoplasma* spp. (e.g. *M. pneumoniae*).

This list of bacteria thus entails bacterial families and species which are involved in pathology of a large number of diseases in humans. Of these, *E. coli* and *B. subtilis* are also used for industrial fermentation; this is also the case for lactic acid bacteria, notably *Lactococcus* spp. and *Lactobacillus* spp. and therefore it is also preferred that the inventive methods, when employed in bacterial systems, are performed on such non-pathogenic species.

When using the inventive methods in a prokaryotic system, it is very often interesting to identify those cells which are impaired in growth or lethally damaged due to the presence of the heterologous expression product introduced according to the invention. This, however, is not completely unproblematic since

the main phenotypic trait associated with e.g. bacterial death is *absence* of the bacterium.

It is therefore necessary to device an experimental setup which will allow identification of cells transformed so as to be less good survivors than other otherwise corresponding cells. One advantage is that expression of the heterologous genetic material is under the control of an inducible promoter. In this way it is possible to expand colonies of cells which have been transformed with genetic material which, when expressed, is lethal or growth-impairing to the cells. After that, the expression can be switched on and careful examination of expanded colonies should reveal those clonal colonies which do not follow the same growth pattern as e.g. an untransformed control.

One method of doing this is to spread transformed cells on plates with growth medium and allow the cells to grow up to a pre-determined average size. The spreading of cells should be such that the visible colonies forming will generally be comprised on one single clone of cells. When the pre-determined size of colonies have been reached, all plates are blotted to a carrier medium so as to prepare a "negative" of each agar plate. After this, the expression of the inserted random sequences is induced and the colonies are allowed to grow again. The plates are examined continuously or at suitable intervals (e.g. by means of digital image processing systems well known to the skilled person). Those colonies which reveal an impaired or arrested growth compared to the remainder of the colonies or compared to controls are thereafter identified, since the growth pattern of each colony in an automated manner can be followed. These colonies can then be identified and isolated from the "negative" blot and it is thereafter a relatively simple procedure to extract the transforming genetic material and determine the sequence thereof.

In this context, one interesting option is to render antibiotic-resistant bacteria non-resistant. The growth medium either contains, or is during culturing enriched with, the antibiotic in question and the colonies which upon induction of expression can be demonstrated to be less drug resistant than controls are examined further. Also pure bactericidal/bacteriostatic effects can be examined. In such an embodiment, the bacteria are e.g. cultured on a suitable growth medium. Those colonies which after induction of expression shows evidence of reduced or arrested growth are examined further: It is expected that some of the bacterial cells will be demonstrated to carry genetic material encoding an expression product which interacts with novel (or known) targets for antibacterial agents.

Another phenotypic trait of interest is of course superior *survival* of cells. It is, when dealing with utilisable bacteria, of interest to identify targets which will increase the survival rate of the bacteria. For example, bacteria used in industrial fermentations normally can be lethally damaged as a consequence of their own uncontrolled production of heterologous expression products. If genes or target molecules can be identified which have a positive effect on the survival of such bacteria, the economic potential is enormous, since a fermentation process will be rendered more economic (less need to startup of new fermentations). Similarly, bacteria used in e.g. waste water purification can be made more resistant against toxic agents in their environment.

The experimental setup in this context is relatively simple: The transformed bacteria are simply subjected to the potentially lethal condition, and only colonies which exhibit a superior survival are isolated and examined (and that will typically be the colonies which are detectable). A setup like the above-described for identifying cell death should thus not be necessary.

Finally, a large group of phenotypic traits to be examined are those which can be detected by e.g. biochemical or immunological means. It is expected that the method of the invention will allow for identification of systems in bacterial cells which, when properly modulated, can render the bacteria more effective as producers in industrial fermentation. Such phenotypic traits could e.g. be changes in enzyme activity, changes in receptor density, changes in expression rate etc.

Special considerations apply when the randomized expression product is fused to a fusion partner which decides the final location of the expression product. Signal sequences in prokaryotes are well-known in the art, but it should briefly be mentioned that membrane-anchoring signals are known, and it is also possible to export the expression product to the periplasmic space of bacteria. Finally, it is also possible to include secretion signals so as to allow the isolation of the expression product from culture supernatant. However, in many cases it is of course most relevant to keep the expression product inside the prokaryotic cytoplasm.

<u>Use of the method in eukaryotic systems</u>

It is especially preferred that the inventive method utilises eukaryotic cells as the substantially identical cells in order to allow screening for active biomolecules. Hence, these eukaryotic cells can be fungal cells, protozoan cells, animal cells, and plant cells.

As is the case for bacteria, a number of fungi are pathogens in mammals, and therefore the present technology will, in parallel with what has been described above concerning antibacterial agents, be useful for identifying antifungal agents by using pathogenic fungi as the substantial identical cells in the method. Furthermore, fungi (especially yeast strains), like

bacteria, are also utilised in fermentation processes (e.g. in the wine and brewing industries), and the method can therefore also be utilised using such non-pathogenic fungi as the substantially identical cells which are transformed with the vectors, whereby improvements in these strains can be obtained.

Preferred examples of fungi serving as the eukaryotic cell in the inventive methods are *Epidermophyton* spp., *Trichophyton* spp., *Microsporum* spp., *Candida albicans*, *Philophora* spp., *Coccidioides immitis*, *Histoplasma capsulatum*, *Blastomyces dermatitidis*, *Paracoccidioides brasiliensis*, *Cryptococcus neoformans*, *Aspergillus* spp., *Saccharomyces cerevisiae*, *Klyveromyces lactis*, and *Piccia pastoris*.

Unlike fungal cells, protozoan cells are only relevant as pathogens for humans and other mammals, although some protozoans form part of biocultures conducting biological waste water purification. The method of the invention is therefore contemplated to be useful in identifying new targets for antiprotozoan agents. In this context, the preferred protozoan cells used as the substantially identical cell in the methods of the invention are selected from the group consisting of *Giardia lamblia*, *Trichomonas vaginalis*, *Dientamoeba fragilis*, *Trypanosoma* spp., *Leishmania* spp., *Entamoeba histolytica*, *Naegleria fowleri*, *Acanthamoeba castellani*, *Harmanella* spp., *Isospora belli*, *Cryptosporidium* spp., *Sarcocystis* spp., *Toxoplasma gondii*, *Plasmodium* spp. (e.g. *P. falciparum*, *P. vivax*, *P. malariae*, *P. knowlesi*, and *P. ovale*), *Babesia* spp., and *Balantidium coli*.

Also plant cells are according to the invention interesting as target eukaryotic cells. The plant cells can be any plant cell which can be subjected to genetic engineering techniques allowing for single cell expression and growth. Thus, cells derived from e.g. *Nicotiana tabacum* (tobacco plant), *Arabidopsis thaliana*, *Brassica napus*, *Brassica juncea*, *Musa sp.* (banana plants),

rice, and corn are examples of plant cells useful in the invention. The skilled person in the art of plant genetic engineering will know to choose suitable plant cells in the appropriate stage of their life cycle, suitable vector systems as well as suitable transformation methods. A short summary is given here:

As other organisms, plant cells can be transformed with foreign DNA. One strategy for plant transformation employs *Agrobacterium tumefaciens*, a naturally occurring plant pathogenic bacterium which contains a plasmid (the Ti plasmid) with the ability to enter plant cells and insert a portion of its genome into plant chromosomes. The Ti plasmid has been engineered to make it a vector for plant transformation by including sequences for replication in *E. coli* and *Agrobacterium*, unique restriction sites for inserting foreign genes, and selectable markers.

Unfortunately, *Agrobacterium* is not very effective at transforming monocots, a large group of plants that includes all of the agriculturally important cereals. Legumes, another important group of food crops, are also difficult to transform and regenerate with *Agrobacterium*. Therefore, a second strategy for transforming plants has been developed involving the Gene Gun, where a gene is inserted into an expression vector and coated onto beads. The DNA-coated beads are then introduced by means of the gene gun into plant cells, where a small fraction are taken up and incorporated into the DNA. Individual plant cells, callus, regenerating shoots, and embryos are all suitable targets in this technique.

To determine whether cells have actually incorporated foreign DNA and become transgenic, reporter genes such as GUS and Luciferase genes are used. In any case, foreign

genes must be flanked by a plant promoter in order to be expressed.

It is preferred to use cells derived from animals. These cells can be mammalian cells, arthropod cells such as insect cells, avian cells, and piscine cells. A number of reasons can be listed for using such cell types which each require a relevant choice of transformation and expression systems, growth conditions, etc, all easily determined and chosen by the skilled person. It suffices to note that e.g. certain insects cause enormous problems in human society (due to their direct damaging activities or due to their functions as vectors carrying infectious agents), and therefore the method of the invention would supplement in the attempts of controlling such insects. As for the mammals, birds and fish, a number of these are important in agri- and aquaculture, where disease control is of interest.

According to the invention mammalian cells such as human cells/cell clones or human cell lines are most preferred. This is due to the fact that a large number of diseases in humans and other mammals etiologically depend on molecular interactions in the living cell - the provision of drugs or lead compounds which interact *in vivo* with biomolecules which play a role in diseases is therefore of great interest. Preferred mammalian cells are Chinese hamster ovary (CHO) cells, VERO cells, HeLa cells, W138 cells, BHK cells, COS-7 293 cells, and MDCK cells, which are all well-known in the art.

The vector is preferably selected from the group consisting of a retroviral vector, a vaccinia virus vector, an adenoviral vector, an adeno associated virus (AAV) vector, a herpes simplex virus (HSV) vector, an alpha virus vector, and a semliki forest virus vector.

The candidate nucleic acids are hence introduced into eukaryotic cells as part of a vector to screen for modulators of target enzyme activity. By the term "introduced into" is herein meant that the nucleic acids enter the cells in a manner suitable for subsequent expression of the nucleic acid. The method of introduction is largely dictated by the targeted cell type, cf. below. Exemplary, but non-limiting methods include $CaPO_4$ precipitation, liposome fusion, lipofectin®, electroporation, viral infection, etc.

The randomly modified nucleic acids are preferably integrated into the host cell genome (e.g. by means of retroviral infection of the host cell), or may exist either transiently or stably in the cytoplasm (*i.e.* through the use of traditional plasmids utilizing standard regulatory sequences, selection markers, etc.). As many pharmaceutically important screens require human or model mammalian cell targets, retroviral vectors capable of transfecting such targets are preferred.

In a preferred embodiment, the candidate nucleic acids are part of a retroviral virion which infects the cells. Generally, infection of the cells is straightforward with the application of the infection-enhancing reagent polybene. Infection can be optimized such that the cells predominantly express a single construct each, e.g. by using the ratio of virus particles to the number of cells. Alternatively, it is possible to "screen out" cells which have been infected with more than one single virion, e.g. by quantitatively assessing a selection marker and only. The rate of infection is well-known to follow a Poisson distribution.

A preferred embodiment of the invention where the substantially identical cells are eukaryotic thus comprises that step (a) is carried out by

(i) transfecting suitable packaging cells with vectors comprising the randomly modified nucleotide sequences and which are integratable in virions produced by said packaging cells,

(ii) culturing said transfected packaging cells in a culture medium under conditions which facilitates production by the packaging cells of virions containing the randomly modified nucleotide sequences,

(iii) recovering and optionally concentrating said virions, and

(iv) transducing said substantially identical cells with the virions.

Thus, the candidate nucleic acids are introduced into the cells using retroviral vectors. Currently, the most efficient gene transfer methodologies harness the capacity of engineered viruses, such as retroviruses, to bypass natural cellular barriers to exogenous nucleic acid uptake. The use is well-known in the art of helper-defective packaging cell-lines which are capable of producing all necessary *trans* proteins (*gag*, *pol*, and *env*) required for packaging, processing, reverse transcription, and integration of recombinant genomes. Those RNA molecules which have in *cis* a ψ packaging signal are packaged into maturing virions. In eukaryotes, retroviruses are preferred for a number of reasons. First, their derivation is fairly easy. Second, unlike Adenovirus-mediated gene delivery, expression from retroviruses is long-term (adenoviruses do not integrate). Adeno-associated viruses have limited space for genes and regulatory units and there is some controversy as to their ability to integrate. Retroviruses therefore currently provide the best compromise in terms of long-term expression, genomic flexibility and stable integration, among other features. The main advantage of retroviruses is that their integration into the host genome allows for their stable transmission through cell division. This ensures that in cell types which undergo multiple independent maturation steps, such as hematopoietic cell progression, the

retrovirus construct will remain resident and continue to express.

Preferred retroviral vectors are derived from retrovirus selected from the group consisting of Avian Leukosis-Sarcoma Virus (ASLV), Mammalian type C, Mammalian type B, and Lentivirus. They may all optionally be modified with heterologous cis-acting elements.

It is preferred that the retroviral vector has non-identical ends so as to facilitate PCR-based generation of random DNA sequences. It is furthermore preferred that these non-identical ends contain non-identical promoters.

An especially preferred retroviral vector contains a heterologous promoter replacing the viral promoter in the 5'-LTR, such as a CMV promoter, an RSV promoter, an SV-40 promoter, a TK promoter, an MT promoter, or an inducible system such as Tet or Ecdysone. A suitable vector is described in WO 96/38553, cf. Fig. 1 and accompanying text.

Particularly well suited retroviral transfection systems are PE501 (US 4,861,719), Bosc23 (WO 94/19478), Psi2 (R. Mulligan/D. Baltimore), GP+E86 (US 5,278,056), PhoenixEco (WO 97/27212), PA317 (US 4,861,719), GP+AM12 (US 5,278,056), DA(ampho) (WO 95/10601, WO 92/05266), Bing (WO 94/19478), FLYA13 (WO 97/08330), ProPak (available from SyStemix), CRIP (R. Mulligan), PsiAM (R. Mulligan/D. Baltimore), Phoenix-Ampho (WO 97/27212), PG13 (Targeted Genetics), H9 (293GPG) (D. Ory, M Sadelain, R. Mulligan, J. Schaffer).

In one embodiment of the present invention, the library is generated in a retrovirus DNA construct backbone, as is generally described in e.g. WO 97/27212. Standard oligonucleotide synthesis is performed to generate the random portion of the

candidate modulator, using techniques well known in the art (cf. Eckstein, Oligonucleotides and Analogues, A Practical Approach, IRL Press At Oxford University Press, 1991); libraries may be commercially purchased. Libraries with up to 10^9 unique sequences can be readily generated in such DNA backbones. After generation of the DNA library, the library is cloned into a first primer. The first primer serves as a "cassette" which is inserted into the retroviral construct. The first primer generally contains a number of elements, including for example, the required regulatory sequences (e.g. translation, transcription, promoters, etc), fusion partners, restriction endonuclease (cloning and subcloning) sites, stop codons (preferably in all three reading frames, regions of complementarity for second strand priming (preferably at the end of the stop codon region as minor deletions or insertions may occur in the random region), etc.

A second primer is then added, which generally consists of some or all of the complementarity region to prime the first primer and optional necessary sequences for a second unique restriction site for subcloning. DNA polymerase is added to make double-stranded oligonucleotides. The double-stranded oligonucleotides are cleaved with the appropriate subcloning restriction endonucleases and subcloned into the target retroviral vectors, described below.

Any number of suitable retroviral vectors may be used. Generally, the retroviral vectors may include: selectable marker genes under the control of internal ribosome entry sites (IRES), which allows for bicistronic operons and thus greatly facilitates the selection of cells expressing peptides at uniformly high levels; and promoters driving expression of a second gene, placed in sense or anti-sense relative to the 5'-LTR (long terminal repeat). Suitable selection genes included, but are not limited to, neomycin, blastocidin, bleomycin, puromycin, and

hygromycin resistance genes, as well as self-fluorescent markers such as green fluorescent protein, enzymatic markers such as lacZ, and surface proteins such as CD8, etc.

Preferred vectors include vectors derived from retrovirus selected from the group consisting of Avian Leukosis-Sarcoma Virus (ASLV), Mammalian type C, Mammalian type B, and Lentivirus as well as vectors derived from MSCV (murine stem cell virus), modified MFG virus and pBABE, and optionally modified with heterologous cis-acting elements.

The retroviruses may include inducible constitutive promoters. For example, there are situations wherein it is necessary to induce peptide expression only during certain phases of the selection process. For instance, a scheme to provide pro-inflammatory cytokines in certain instances must include induced expression of the peptides. This is because there is some expectation that over-expressed pro-inflammatory drugs might in the long-term be detrimental to cell growth. Accordingly, constitutive expression is undesirable, and the peptide in only turned on during that phase of the selection process when the phenotype is required, and then the peptide is shut down by turning off the retroviral expression to confirm the effect or ensure long-term survival of the producer cells. A large number of both inducible and constitutive promoters are known to the skilled person.

In addition, it is possible to configure a retroviral vector to allow inducible expression of retroviral inserts after integration of a single vector in target cells; importantly, the entire system is contained within the single retrovirus. Tet-inducible retroviruses have been designed incorporating the Self-Inactivating (SIN) feature of 3'LTR enhancer/promoter retroviral deletion mutant (Hoffmann *et al.*, PNAS U.S.A. 93:5185 (1996)). Expression of this vector in cells is virtually undetectable in

the presence of tetracycline or other active analogues. However, in the absence of Tet, expression is turned on to maximum within 48 hours after induction, with uniform increased expression of the whole population of cells that harbour the inducible retrovirus, indicating that expression is regulated uniformly within the infected cell population. a similar, related system uses a mutated Tet DNA-binding domain such that it bound DNA in the presence of Tet, and was removed in the absence of Tet. Either of these systems is suitable.

In this manner the primers create a library of fragments, each containing a different random nucleotide sequence within a scaffold sequence derived from genetic material encoding a enzyme modulator. The ligation products are then transformed into bacteria, such as *E. coli* and DNA is prepared from the resulting library, as is generally outlined in Kitamura, PNAS U.S.A. 92: 9146-50 (1995), which is incorporated by reference herein.

Delivery of the library DNA into a retroviral packaging system results in conversion to infectious virus. Suitable retroviral packaging system cell lines include, but are in no way limited to, PE501, Bosc23, Psi2, GP+E86, PhoenixEco, PA317, GP+AM12, DA(ampho), Bing, FLYA13, ProPak, CRIP, PsiAM, Phoenix-Ampho, PG13, and H9 (293GPG).

In a preferred embodiment, the cell lines discussed above, and the other methods for producing retrovirus, are useful for production of virus by transient transfection. The virus can be either used directly or used to infect another retroviral producer cell so as to expand the library.

Concentration of virus may be done as follows: Generally, retroviruses are titred by applying retrovirus-containing supernatant onto indicator cells, such as NIH3T3 cells, and then measuring

the percentage of cells expressing phenotypic consequences of infection. The concentration of the virus is determined by multiplying the percentage of cells infected by the dilution factor involved, and taking into account the number of target cells available to obtain a relative titre. If the retrovirus contains a reporter gene, such as *lacZ*, then infection, integration, and expression of the recombinant virus is measured by histological staining for *lacZ* expression or by flow cytometry (FACS). In general, retroviral titres generated from even the best of the producer cells do not exceed 10^7 per ml, unless concentration is performed on relatively expensive or exotic apparatus. However, it is believed that particles as large as retrovirus will not move very far by means of brownian motion in liquid, fluid dynamics predictions show that much of the virus never comes in contact with the cells in order to initiate infection. However, if cells are grown or placed on a porous filter surface and retrovirus are allowed to pass the cells by gradual gravitometric flow, a high concentration of virus around cells can be effectively maintained at all times. Thus, up to a ten-fold higher infectivity by infecting cells on a porous membrane and allowing retrovirus supernatant to flow past them has been seen. This should allow titres of 10^8 after concentration.

Upon isolation/concentration of virus, the substantially identical target cells are transduced by methods well-known in the art.

As mentioned above, fusion of the expression product to at least one fusion partner which facilitates expression and/or purification/isolation and/or further stabilization of the expression product is often desired.

In eukaryotes, the fusion partner is often a sorting signal or a targeting sequence. Such a sorting signal will be in the form of

a signal patch or a signal peptide. The well-known function of a sorting signal is to effect export of an expressed peptide into endoplasmic reticulum, into Golgi apparatus, into lysosomes, into secretory vesicles, into mitochondria, into peroxisomes, or into the nucleus. Of course, also export to the membrane or out of the cell are possibilities. Preferably, the sorting signal or targeting sequence is selected from the group consisting of

- a nuclear localization signal (NLS) such as Pro-Lys-Lys-Lys-Arg-Lys-Val (SV40 large T antigen NLS), Ala-Arg-Arg-Arg-Arg-Pro (human retinoic acid receptor-β NLS), Glu-Glu-Val-Gln-Arg-Lys-Arg-Gln-Lys-Leu (NFκB p50), Glu-Glu-Lys-Arg-Lys-Arg-Thr-Tyr-Glu (NFκB p65), and Ala-Val-Lys-Arg-Pro-Ala-Ala-Thr-Lys-Lys-Ala-Gly-Gln-Ala-Lys-Lys-Lys-Lys-Leu-Asp (Xenopus nucleoplasmin NLS);
- a membrane anchoring sequence such as those derived from CD8, ICAM-2, IL-8, CD4, and LFA-1, and a lipidation sequence such as a myristylation or a palmitoylation sequence;
- a lysosomal sorting signal such as a lysosomal degradation sequence, and a lysosomal membrane sequence
- a mitochondrial localization sequence such as a mitochondrial matrix sequence, a mitochondrial inner membrane sequence, a mitochondrial intermembrane space sequence, and a mitochondrial outer membrane sequence; an endoplasmic reticulum localization sequence such as the sequence from calreticulin (KDEL) and the sequence from adenovirus E3/19K protein (LULSRRSFIDEKKMP);
- a peroxisome sequence such as the peroxisome matrix sequence from Luciferase;
- a farnesylation sequence such as LNPPDESGPGCMSCKCVLS;
- a geranylgeranylation sequence such as LTEPTQPTRNQCCSN;
- a destruction sequence such as RTALGDIGN; and
- a secretory signal sequence such as the secretory signals from IL-2, growth hormone, preproinsulin, and influenza HA protein.

Enzyme modulators useful in the invention

The art has demonstrated the existence of numerous effective peptide enzyme activity modulators. Especially the enzyme inhibitors are well-characterized. Non-limiting examples which are all incorporated by reference herein are listed in the following:

BPTI/Kunitz family of protease inhibitors:

Pancreatic trypsin inhibitor (BPTI) from *Bos taurus;*
Spleen trypsin inhibitor from *Bos taurus;*
Inter-alpha-trypsin inhibitor light chain (bikunin) from *Bos taurus, Homo sapiens, Meriones unguiculatus, Mesocricetus auratus, Mus musculus, Sus scrofa, Pleuronectes platessa,* and *Rattus norvegicus, respectively;*
Inter-alpha-trypsin inhibitor from *Equus caballus, Ovis aries,* and *Capra hircus,* respectively;
Hemolymph trypsin inhibitor A from *Manduca sexta;*
Hemolymph trypsin inhibitor B from *Manduca sexta;*
Colostrum trypsin inhibitor from *Bos taurus;*
Trypstatin from *Rattus norvegicus;*
Proteinase inhibitor from *Tachypleus tridentatus;*
Serum basic protease inhibitor from *Bos taurus;*
Chymotrypsin inhibitor SCI-III from *Bombyx mori;*
Male accessory gland serine-protease inhibitor from *Drosophila funebris;*
Protease inhibitor 5 II from *Anemonia sulcata;*
Chymotrypsin inhibitors SCI-I and SCI-II from *Bombyx mori;*
Proteinase inhibitors SHPI and SHPI-2 from *Stoichactis helianthus;*
Isoinhibitor K from *Helix pomatia;*
Trypsin inhibitor IV from *Radianthus macrodactylus;*

Venom basic protease inhibitors IX and VIIIB from *Bungarus fasciatus;*

Venom basic protease inhibitors I and III from *Vipera ammodytes ammodytes;*

Venom basic protease inhibitor II from *Daboia russelli siamensis, Hemachatus haemachatus,* and *Naja nivea,* respectively;

Venom basic protease inhibitors B and E from *Dendroaspis polylepis polylepis;*

Venom chymotrypsin inhibitor from *Naja naja;*

Venom basic protease inhibitors I and K from *Dendroaspis polylepis polylepis;*

Venom basic protease inhibitor K from *Dendroaspis angusticeps*

Venom trypsin inhibitor from *Eristocophis macmahoni* and *Naja naja,* respectively;

Protease inhibitor from *Sarcophaga bullata;*

Tissue factor pathway inhibitor from *Homo sapiens, Oryctolagus cuniculus,* and *Rattus norvegicus,* respectively;

Tissue factor pathway inhibitor 2 from *Homo sapiens;*

Uterine plasmin/trypsin inhibitor from *Sus scrofa;*

Protease nexin II (fragment of Alzheimer's disease amyloid A4 protein) from *Homo sapiens, Mus musculus, rattus norvegicus, Macaca fascicularis* and *Macaca mulatta,* respectively;

Amyloid protein 2 from *Homo sapiens* and *Rattus norvegicus,* respectively; and

Ornithodorin from *Ornithodoros moubata,*

as well as inhibitors homologous therewith isolated from other sources than those explicitly mentioned.

Serpin family of protease inhibitors:

Alpha-1-proteinase inhibitors (alpha-1-antitrypsins)from *Equus caballus, Mus musculus, Cavia porcellus, Oryctolagus cuniculus, Bos taurus, Chinchilla villidera, Didelphis marsurpiales virginiana, Homo sapiens, Macropus eugenii, Mus caroli, Papio*

anubis, *Sus scrofa*, *Rattus norvegicus*, and *Ovis aries*, respectively;

Alpha-1-antichymotrypsin from *Homo sapiens*;

Antithrombin III from *Homo sapiens*, *Bos taurus*, *Mus musculus*, *Ovis aries*, *Mesocricetus auratus*, and *Gallus gallus*, respectively;

Alpha-2-plasmin inhibitor (alpha-2-antiplasmin) from *Bos taurus*, *Homo sapiens*, and *Mus musculus*, respectively;

Bomapin (Protease Inhibitor 10) from *Homo sapiens*;

Contrapsin from *Mus musculus* and *Cavia porcellus*;

Contrapsin-like protease inhibitors from *Rattus norvegicus*;

Factor XIIa inhibitor from *Bos taurus*;

Glia derived nexin (protease nexin I) from *Homo sapiens*, *Mus musculus*, and *Rattus norvegicus*, respectively;

Heparin co-factor II from *Homo sapiens*, *Mus musculus*, *Oryctolagus cuniculus*, and *Rattus norvegicus*, respectively;

47 kDa heat shock protein (serine protease inhibitor J6) from *Mus musculus* and *Gallus gallus*, respectively;

C1-inhibitor from *Homo sapiens*;

Leukocyte elastase inhibitor from *Equus caballus*, *Homo sapiens*, and *Sus scrofa*, respectively;

Protein C inhibitor from *Homo sapiens*;

Kallistatin from *Homo sapiens*;

Kallikrein-binding protein from *Mus musculus* and *Rattus norvegicus*, respectively;

Maspin from *Homo sapiens*, *Mus musculus*, and *Rattus norvegicus*, respectively;

Plasminogen activator inhibitor-1 from *Bos taurus*, *Homo sapiens*, *Mus musculus*, *Mustela vison*, and *Rattus norvegicus*, respectively;

Plasminogen activator inhibitor-2 from *Homo sapiens*, *Mus musculus*, and *Rattus norvegicus*, respectively;

Neuroserpin from *Homo sapiens*, *Mus musculus*, and *Gallus gallus*

Cytoplasmic antiproteinases 1, 2 and 3 from *Homo sapiens*

Antitrypsin from *Bombyx mori*, respectively;

Antichymotrypsins I and II from *Bombyx mori*;

Alaserpin from *Manduca sexta*;

Serine protease inhibitor 2.1 from *Rattus norvegicus*;

Serine proteinase inhibitor 1 from Cowpox virus, Rabbit pox virus, Swine pox virus, Vaccinia virus, and Variola virus, respectively;

Serine proteinase inhibitor 2 from Rabbit pox virus, Vaccinia virus, and Variola virus, respectively;

Serine proteinase inhibitor 3 from Vaccinia virus and Variola virus, respectively;

Serpin 1 from Myxoma virus;

Serine proteinase inhibitor from *Halocynthia roretzi*;

Ice inhibitor from Cowpox virus (thiol-protease inhibitor);

as well as inhibitors homologous therewith isolated from other sources than those explicitly mentioned.

Kazal family of protease inhibitors:

Acrosin inhibitors from *Bos taurus*, *Homo sapiens*, *Sus scrofa*, and *Macaca fascicularis*, respectively;

Elastase inhibitor from *Anemonia sulcata*;

Ovoinhibitor from *Gallus gallus*;

Rhodniin from *Rhodnius prolixus*;

Pancreatic secretory trypsin inhibitor from *Rattus norvegicus*, *Anguilla anguilla*, *Bos taurus*, *Canis familiaris*, *Homo sapiens*, *Sus scrofa* and *Ovis aries*, respectively;

Pancreatic secretory trypsin inhibitor II from *Rattus norvegicus*;

Double-headed protease inhibitor (submandibular gland) from *Canis familiaris*, *Felis silvestris catus*, *Meles meles*, *Panthera leo*, and *Vulpes vulpes*, respectively;

Trypsin inhibitor from *Halocynthia roretzi*;

Tryptase inhibitor from *Hirudo medicinalis*;

Prostatic secretory glycoprotein from *Mus musculus* ; Ovomucoid (third domain) from *Aburria pipile, Aepypodius arfakianus, Afropavo congensis, Alectoris chukar, Alectoris rufa, Anas platyrhynchos, Chloephaga picta, Cyanochen cyanoptera, Neochen jubata, Tadorna radjah, Lophonetta specularioides, Anas capensis, Aix galericulata, Aix sponsa, Sarkidiornis melanotos, Alopochen aegyptiaca, Mergus cucullatus, Anhinga novaehollandiae, Anser anser anser, Anser indicus, Anseranas semipalmata, Arborophila torqueola, Argusianus argus, Aythya americana, Netta rufina, Balearica pavonina, Bambusicola thoracica, Bonasa umbellus, Branta canadensis, Anser canagicus, Callipepla squamata castanogastric, Callipepla squamata pallida, Carpodacus mexicanus, Carpococcyx renauldi, Casuarius casuarius, Casuarius bennetti, Cereopsis novaehollandiae, Chauna chavaria, Chauna torquata, Gallus gallus, Chrysolophus amherstiae, Chrysolophus pictus, Circus aeruginosus, Colinus virginianus, Corvus albus, Corvus monedula, Coscoroba coscoroba, Coturnix delegorguei, Coturnix coturnix japonica, Crossoptilon crossoptilon, Cygnus atratus, Cygnus olor, Oxyura jamaicensis, Oxyura vittata, Cyrtonyx montezumae, Dacelo novaeguineae, Dendrocygna arborea, Dendrocygna arcuata, Dendrocygna autumnalis, Dendrocygna bicolor, Dendrocygna eytoni, Dendrocygna viduata, Dromaius novae-hollandiae, Eudromia elegans, Francolinus afer coqui, Francolinus erckelii, Francolinus francolinus, Francolinus pondicerianus, Fulica atra, Gallinula chloropus, Gallirallus australis, Gallus varius, Geococcyx californianus, Grus carunculatus, Grus japonensis, Grus vipio, Anthropoides virgo, Guira guira, Guttera pucherani, Gyps coprotheres, Polyboroides radiatus, Aquila audax, Necrosyrtes monachus, Haliaeetus albicilla, Haliastur indus, Larus ridibundus, Larus marinus, Vanellus spinosus, Leipoa ocellata, Lophura bulweri, Lophortyx californica, Lophortyx gambelii, Lophura ignita, Lophura diardi, Lophura leucomelana, Megapodius freycinet, Meleagris gallopavo, Agriocharis ocellata, Nothoprocta cinerascens, Nothoprocta*

perdicaria, Numida meleagris, Acryllium vulturinum, Nycticorax nycticorax, Opisthocomus hoazin, Oreortyx pictus, Ortalis vetula, Pavo cristatus, Pavo muticus, Penelope jacquacu, Penelope superciliaris, Perdrix perdrix, Phasianus colchicus colchicus, Phasianus versicolor, Phalacrocorax sulcirostris, Podargus strigoides, Polyplectron bicalcaratum, Polyplectron emphanum, Polyborus plancus, Pygoscelis adelie, Phalacrocorax albiventer, Rhea americana, Pterocnemia pennata, Rhynchotus rufescens, Rollulus roulroul, Scythrops novaehollandiae, Spheniscus humboldti, Struthio camelus, Syrmaticus mikado, Syrmaticus reevesii, Tinamus major, Turdus merula, Turnix sylvatica, Tympanuchus cupido, Centrocercus urophasianus, Tragopan blythii, Tragopan caboti, Tragopan satyra, Tragopan temminckii, Lophophorus impejanus, Crossoptilon auritum, Crossoptilon mantchuricum, Lophura edwardsi, Lophura nycthemera, Lophura swinhoei, Pucrasia macrolopha, Catreus wallichii, Syrmaticus ellioti, Syrmaticus humiae, Syrmaticus soemmerringii, Lagopus leucurus, and *Vultur gryphus*, respectively,

as well as inhibitors homologous therewith isolated from other sources than those explicitly mentioned.

Soybean trypsin inhibitor (Kunitz) family of protease inhibitors:

Aspartic proteinase inhibitor from *Solanum tuberosum;*
Cathepsin D inhibitors from *Solanum tuberosum;*
Wound-induced aspartate proteinase inhibitor from *Solanum tuberosum;*
Chymotrypsin inhibitor 3 precursor from *Psophocarpus tetragonolobus;*
Trypsin inhibitor from *Adenanthera pavonina;*
Trypsin inhibitor from *Prosopsis juliflora;*
Trypsin inhibitor from *Erythrina caffra;*
Trypsin inhibitor from *Erythrina latissima;*

Chymotrypsin inhibitor from *Erythrina variegata*;

Trypsin inhibitors 1A, 1B, and 2 from *Psophocarpus tetragonolobus*;

Trypsin/chymotrypsin inhibitor from *Alocasia macrorrhiza*;

Trypsin inhibitor from *Albizzia julibrissin*;

Trypsin inhibitor from *Acacia confusa*;

Trypsin inhibitors A, B and C from *Glycine max*;

Trypsin inhibitors KTI1and KTI2 from *Glycine max*;

Latex serine proteinase inhibitor from *Carica papaya*;

Cysteine proteinase inhibitor PCPI 8.3 from *Solanum tuberosum*;

Kunitz-type inhibitors 1 and 2 (PKI-1) from *Solanum tuberosum*;

Potato inhibitor I family (protease inhibitors):

Trypsin/subtilisin inhibitor from *Amaranthus caudatus*;

Subtilisin inhibitor from *Momordica charantia*;

Wound-induced proteinase inhibitor I from *Lycopersicon esculentum*, *Lycopersicon peruvianum*, and *Solanum tuberosum*, respectively;

Subtilisin inhibitors I and II from *Phaseolus angularis*;

Proteinase inhibitor I from *Solanum tuberosum*;

Chymotrypsin inhibitor 2A (CI-2A) from *Hordeum vulgare*;

Chymotrypsin inhibitor 2B (CI-2B) from *Hordeum vulgare*;

Chymotrypsin inhibitor 1A (CI-1A) from *Hordeum vulgare*;

Chymotrypsin inhibitor 1B (CI-1B) from *Hordeum vulgare*;

Chymotrypsin inhibitor 1C (CI-1C) from *Hordeum vulgare*;

Chymotrypsin inhibitor I, a, b and c subunits from *Solanum tuberosum*;

Subtilisin inhibitor from *Vicia faba*;

Ethylene-responsive proteinase inhibitor from *Lycopersicon esculentum*;

Proteinase inhibitors I-A and I-B from *Nicotiana tabacum*;

Eglin C from *Hirudo medicinalis*;

Inhibitor of trypsin and Hageman factor from *Cucurbita maxima*;

Trypsin inhibitor MCI-3 from *Momordica charantia*; and
Trypsin inhibitor I from *Nicotiana sylvestris*,
as well as inhibitors homologous therewith isolated from other sources than those explicitly mentioned.

Bowman-Birk family of protease inhibitors:

Bowman-birk type protease inhibitors from *Arachis hypogaea*, *Coix lachryma-jobi*, *Phaseolus angularis*, *Glycine max*, *Triticum aestivum*, *Arachis hypogaea*, *Phaseolus vulgaris*, *Setaria italica*, *Dolichos axillaris*, *Lonchocarpus capassa*, *Oryza sativa*, *Hordeum vulgare*, *Medicago scutellata*, *Phaseolus aureus*, *Phaseolus lunatus*, *Vicia angustifolia*, *Vicia faba*, and *Vigna unguiculata*, respectively; and
Wound induced trypsin inhibitors from *Medicago sativa* and *Zea mays*, respectively,
as well as inhibitors homologous therewith isolated from other sources than those explicitly mentioned.

Squash inhibitor family:

Elastase inhibitor from *Momordica charantia*;
Trypsin inhibitors I, II, and III from *Lagenaria leucantha*;
Trypsin inhibitor I from *Citrullus vulgaris*;
Trypsin inhibitors I, III, and IV from *Cucurbita maxima*;
Trypsin inhibitors I and II from *Luffa cylindrica*;
Trypsin inhibitors I and II from *Momordica charantia*;
Trypsin inhibitor I from *Momordica repens*;
Trypsin inhibitor II from *Bryonia dioica*;
Trypsin inhibitors II3 and IV from *Cucumis sativus*;
Trypsin inhibitor II from *Ecballium elaterium*;
Trypsin inhibitors I, II and III from *Cucumis melo* var. *Conomon*;
Trypsin inhibitors II and III from *Cucurbita pepo*; and
Trypsin inhibitor A from *Momordica charantia*,

as well as inhibitors homologous therewith isolated from other sources than those explicitly mentioned.

***Streptomyces* subtilisin inhibitor (SSI) family:**

Subtilisin inhibitor from *Streptomyces albogriseolus*;
Subtilisin inhibitor-like protein-1 from *Streptomyces cacaoi*;
Trypsin inhibitor STI2 from *Streptomyces longisporus*;
Subtilisin inhibitor-like protein-2 from *Streptomyces rochei*;
Subtilisin inhibitor-like protein-3 from *Streptomyces coelicolor*;
Subtilisin inhibitor-like protein-4 from *Streptomyces lavendulae*;
Subtilisin inhibitor-like protein-3 from *Streptomyces virginiae*;
Protease inhibitor SIL-V3 from *Streptoverticillium eurocidicus*;
Protease inhibitor SIL-V1/SIL-V4 from *Streptoverticillium flavopersicus*;
Protease inhibitor SIL-V5 from *Streptoverticillium luteoverticillatus*;
Protease inhibitor SIL-V2 from *Streptoverticillium orinoci*;
Alkaline protease inhibitor 2C' from *Streptomyces griseoincarnatus*;
Protease inhibitor from *Streptomyces lividans*; and
Plasminostreptin from *Streptomyces antifibrinolyticus*,
as well as inhibitors homologous therewith isolated from other sources than those explicitly mentioned.

***Bombyx* family of protease inhibitors**

Fungal protease inhibitor F (FPI-F) from *Bombyx mori*, as well as inhibitors homologous therewith isolated from other sources than *Bombyx mori*.

Wap-type 'Four-disulphide Core' Proteinase inhibitors

Antileukoproteinase 1 from *Homo sapiens* and *Mus musculus*, respectively;
Elafin from *Homo sapiens* and *Sus scrofa*, respectively; and
Chelonianin (basic protease inhibitor) from red sea turtle, as well as inhibitors homologous therewith isolated from other sources than those explicitly mentioned.

Hirudin family of protease inhibitors

Hirudins from *Hirudo medicinalis* and *Hirudinaria manillensis*, respectively, as well as inhibitors homologous therewith isolated from other sources than those explicitly mentioned.

Factor Xa inhibitors

Antistasin from *Haementeria officinalis, Haementeria ghilianii, Hirudo medicinalis, Hirudo nipponia*, and *Hydra magnipapillata*, respectively, as well as inhibitors homologous therewith isolated from other sources than those explicitly mentioned.

Ascaris trypsin inhibitor family

Chymotrypsin/elastase inhibitors from *Ascaris suum* and
Trypsin inhibitor from *Ascaris suum* as well as inhibitors homologous therewith isolated from other sources than *Ascaris suum*.

Cystatin family of protease inhibitors

Leukocyte cysteine proteinase inhibitors 1 and 2 from *Sus scrofa*;
Stefins 1, 2 and 3 from *Mus musculus*;
Cystatins A1, A5, A8 and B from *Sus scrofa*;

Cystatins A and B from *Bos taurus*;

Stefin C from *Bos taurus*;

Cystatin B from *Ovis aries*;

Cystatins A, C, D, M, SN, S and SA from *Homo sapiens*;

Cystatins B and C from *Mus musculus*;

Cystatins C and S from *Rattus norvegicus*;

Cystatins C from *Macaca mulatta* and *Saimiri sciureus*, respectively;

Cystatin alpha (epidermal thiol proteinase inhibitor) from *Rattus norvegicus*;

Cystatin B (liver thiol proteinase inhibitor) from *Homo sapiens* and *Rattus norvegicus*, respectively;

Cystatin (colostrum thiol proteinase inhibitor) from *Bos taurus*;

Cystatins from *Gallus gallus* and *Coturnix coturnix japonica*, respectively;

Cystatin from *Cyprinus carpio*;

Cystatin from *Bitis arientas*;

Cystatin I from *Zea mays*;

Oryzacystatin-I from *Oryza sativa*;

Oryzacystatin-II from *Oryza sativa*;

Cystatin A from *Helianthus annuus*;

Cystatin B from *Helianthus annuus*;

Cystatin from *Vigna unguiculata*;

Onchocystatin from *Onchocerca volvulus*;

Cysteine proteinase inhibitor from *Solanum tuberosum*;

Cystatin WCPI-3 from *Wisteria floribunda*;

Cystatin from *Glycine max*; and

Kininogens from *Bos taurus, Homo sapiens* and *Rattus norvegicus*, respectively,

as well as inhibitors homologous therewith isolated from other sources than those explicitly mentioned.

Calpain family of cysteine protease inhibitors

Calpastatin from *Bos taurus, Cercopithecus aethiops, Homo sapiens, Mus musculus, Sus scrofa, Oryctolagus cuniculus, Rattus norvegicus*, and *Ovis aries*, respectively, as well as inhibitors homologous therewith isolated from other sources than those explicitly mentioned.

Tissue inhibitor of metalloproteinases family

Metalloproteinase inhibitor 1 from *Bos taurus, Homo sapiens, Mus musculus, Papio cynocephalus, Sus scrofa, Oryctolagus cuniculus, Rattus norvegicus*, and *Ovis aries*, respectively;
Metalloproteinase inhibitor 2 from *Bos taurus, Homo sapiens, Mus musculus, Rattus norvegicus*, and *Gallus gallus*, respectively; and
Metalloproteinase inhibitor 3 from *Bos taurus, Homo sapiens, Mus musculus, Rattus norvegicus*, and *Gallus gallus*, respectively, as well as inhibitors homologous therewith isolated from other sources than those explicitly mentioned.

Carboxypeptidase A inhibitors

Carboxypeptidase A inhibitor from *Ascaris suum*, as well as inhibitors homologous therewith isolated from other sources than *Ascaris suum*.

Metallocarboxypeptidase inhibitors

Metallocarboxypeptidase inhibitors from *Lycopersicon esculentum* and *Solanum tuberosum*, respectively, as well as inhibitors homologous therewith isolated from other sources than those explicitly mentioned.

Angiotensin-converting enzyme inhibitors

Angiotensin-converting enzyme inhibitor from *Thunnus albacares*;
Angiotensin-converting enzyme inhibitors from *Bothrops insularis*;
Angiotensin-converting enzyme inhibitors from *Bothrops jararaca*;
Angiotensin-converting enzyme inhibitor from *Agkistrodon halys blomhoffi*;
Angiotensin-converting enzyme inhibitor from *Agkistrodon halys pallas*; and
Angiotensin-converting enzyme inhibitor from *Vipera aspis*, as well as inhibitors homologous therewith isolated from other sources than those explicitly mentioned.

Protease inhibitors not belonging to a specific family

Ecotin from *Escherichia coli*;
Metalloproteinase inhibitor from *Streptomyces nigrescens*;
Proteinase inhibitor from *Erwinia chrysanthemi*;
Proteinase inhibitor from *Pseudomonas aeruginosa*;
Protease A inhibitor 3 from *Saccharomyces cerevisiae*;
Protease B inhibitors 1 and 2 from *Saccharomyces cerevisiae*;
Major pepsin inhibitor PI-3 from *Ascaris suum*;
Intracellular proteinase inhibitor from *Bacillus subtilis*;
Proteinase inhibitor PTI from *Solanum tuberosum*;
Proteinase inhibitor PCI-I from *Solanum tuberosum*;
Proteinase inhibitor IIA from *Solanum tuberosum*;
Proteinase inhibitor IIB from *Solanum tuberosum*;
Proteinase inhibitors A and B from *Sagittaria sagittifolia*;
Proteinase inhibitor from *Solanum melongena*;
Trypsin inhibitor from *Brassica napus*;
Trypsin inhibitor 2 from *Sinapis alba*;
Trypsin inhibitor from *Zea mays*;
Trypsin inhibitor from *Sinapis arvensis*;

Trypsin inhibitor from *Trichosanthes kirilowii*;
Wound-induced proteinase inhibitor II from *Lycopersicon esculentum*;
Protease inhibitors LCMI I and PMP-D2 from *Locusta migratoria*;
Protease inhibitor from *Bacillus brevis*;
Marinostatins C-1, C-2 and D from *Alteromonas* sp.; and
Host protease inhibitor from bacteriophage T4,
as well as inhibitors homologous therewith isolated from other sources than those explicitly mentioned.

Cereal alpha-amylase/trypsin inhibitor family

Alpha amylase/trypsin inhibitor CM1 from *Triticum aestivum*;
Alpha amylase/trypsin inhibitor CM2 from *Triticum aestivum*;
Alpha amylase/trypsin inhibitor CM3 from *Triticum aestivum*;
Alpha amylase/trypsin inhibitor CM16 from *Triticum aestivum*;
Alpha amylase/trypsin inhibitor CM17 from *Triticum aestivum*;
Alpha amylase inhibitor 0.19 from *Triticum aestivum*;
Alpha amylase inhibitor 0.28 from *Triticum aestivum*;
Alpha amylase inhibitor 0.53 from *Triticum aestivum*;
Alpha amylase inhibitor WDAI-3 from *Triticum aestivum*;
Alpha amylase/trypsin inhibitor CMA from *Hordeum vulgare*;
Alpha amylase/trypsin inhibitor CMB from *Hordeum vulgare*;
Alpha amylase/trypsin inhibitor CMC from *Hordeum vulgare*;
Alpha amylase/trypsin inhibitor CMD from *Hordeum vulgare*;
Alpha amylase inhibitor CME from *Hordeum vulgare*;
Alpha amylase inhibitor BMAI-1 from *Hordeum vulgare*;
Alpha amylase inhibitor BDAI-1 from *Hordeum vulgare*;
Alpha amylase/trypsin inhibitor from *Eleusine coracana*; and
Trypsin/factor XIIa inhibitor from *Zea mays*,
as well as inhibitors homologous therewith isolated from other sources than those explicitly mentioned.

Alpha-amylase/trypsin inhibitor homologous to thaumatin

Alpha-amylase/trypsin inhibitor from *Zea mays* as well as inhibitors homologous therewith isolated from other sources than *Zea mays*.

Alpha-amylase/subtilisin inhibitor family

Alpha-amylase/subtilisin inhibitor from *Hordeum vulgare*;
Alpha-amylase/subtilisin inhibitor from *Triticum aestivum*; and
Alpha-amylase/subtilisin inhibitor from *Oryzae sative*,
as well as inhibitors homologous therewith isolated from other sources than those explicitly mentioned.

Inhibitors of insect alpha-amylases

Small protein inhibitor of insect alpha-amylases 1, 2 and 3 from *Sorghum bicolor milo* as well as inhibitors homologous therewith isolated from other sources than *Sorghum bicolor milo*.

Inhibitors of mammalian alpha-amylases derived from Streptomyces species

Alpha-amylase inhibitor HAIM I from *Streptomyces griseosporeus*;
Alpha-amylase inhibitor PAIM I from *Streptomyces olivaceoviridis*;
Alpha-amylase inhibitor HAIM II from *Streptomyces griseosporeus*;
Alpha-amylase inhibitor PAIM II from *Streptomyces olivaceoviridis*;
Alpha-amylase inhibitor AI-3688 from *Streptomyces aureofaciens*;
Alpha-amylase inhibitor Z-2685 from *Streptomyces rochei*; and
Alpha-amylase inhibitor HOE-467A from *Streptomyces tendae*,
as well as inhibitors homologous therewith isolated from other sources than those explicitly mentioned.

Trehalase inhibitors

Trehalase inhibitor from *Periplaneta americana* as well as inhibitors homologous therewith isolated from other sources.

Polygalacturonase inhibitors

Polygalacturonase inhibitors from *Phaseolus vulgaris* and *Pyrus communis* as well as inhibitors homologous therewith isolated from other sources than those explicitly mentioned.

Fucosyltransferase inhibitors

Fuctinins from *Rattus norvegicus* as well as inhibitors homologous therewith isolated from other sources.

Protein kinase C inhibitors

14-3-3 Protein beta/alpha from *Bos taurus, Ovis aries, Homo sapiens,* and *Rattus norvegicus*, respectively;
14-3-3 Protein epsilon from *Mus musculus, Ovis aries, Homo sapiens,* and *Rattus norvegicus*, respectively;
14-3-3 Protein eta from *Bos taurus, Mus musculus,* and *Rattus norvegicus*, respectively;
14-3-3 Protein gamma from *Bos taurus, Ovis aries,* and *Rattus norvegicus*, respectively;
14-3-3 Protein zeta/delta from *Bos taurus, Mus musculus, Ovis aries, Homo sapiens,* and *Rattus norvegicus*, respectively;
Hint protein from *Bos taurus, Rattus norvegicus, Homo sapiens, Mus musculus,* and *Oryctolagus cuniculus*, respectively; and
14 kDa zinc-binding protein from *Brassica juncea* and *Zea mays*, respectively,
as well as inhibitors homologous therewith isolated from other sources than those explicitly mentioned.

cAMP-dependent protein kinase inhibitors

cAMP-dependent protein kinase inhibitor (muscle/brain form) from *Homo sapiens, Oryctolagus cuniculus, Mus musculus,* and *Rattus norvegicus*, respectively, and cAMP-dependent protein kinase inhibitor (testis form) from *Mus musculus,* and *Rattus norvegicus*, respectively, as well as inhibitors homologous therewith isolated from other sources than those explicitly mentioned.

Cyclic nucleotide phosphodiesterase inhibitor

Cyclic nucleotide phosphodiesterase inhibitor from *Dictyostelium discoideum* as well as inhibitors homologous therewith isolated from other sources.

Protein phosphatase inhibitors

Protein phosphatase inhibitor 1 from *Homo sapiens, Oryctolagus cuniculus,* and *Rattus norvegicus*, respectively;
Protein phosphatase inhibitor 2 from *Homo sapiens, Oryctolagus cuniculus,* and *Rattus norvegicus*, respectively;
Heat-stable protein phosphatase 2a inhibitor I1PP2A from *Bos taurus* and *Homo sapiens*, respectively; and
Phosphatase RAPA inhibitor from *Bacillus subtilis,*
as well as inhibitors homologous therewith isolated from other sources than those explicitly mentioned.

TCD/MRS6 family of GDP dissociation inhibitors

Secretory pathway GDP dissociation inhibitor from *Saccharomyces cerevisiae;*
Rab GDP dissociation inhibitor alpha from *Bos taurus, Homo sapiens, Mus musculus,* and *Rattus norvegicus*, respectively;

Rab GDP dissociation inhibitor beta from *Homo sapiens*, *Mus musculus*, and *Rattus norvegicus*, respectively;
Rho GDP-dissociation inhibitor 1 from *Bos taurus*, *Homo sapiens*, *Caenorhabditis elegans* and *Cavia porcellus*, respectively;
Rho GDP-dissociation inhibitor from *Saccharomyces cerevisiae* ;
Rho GDP-dissociation inhibitor 2 from *Homo sapiens* and *Mus musculus*, respectively; and
Rho GDP-dissociation inhibitor from *Mus musculus*,
as well as inhibitors homologous therewith isolated from other sources than those explicitly mentioned.

ATPase inhibitors

Mitochondrial ATPase inhibitors from *Bos taurus*, *Caenorhabditis elegans*, *Pichia jadinii*, *Sus scrofa*, *Rattus norvegicus*, and *Saccharomyces cerevisiae*, respectively, and
Sodium/potassium ATPase inhibitor from *Sus scrofa*,
as well as inhibitors homologous therewith isolated from other sources than those explicitly mentioned.

Phospholipase A2 inhibitory proteins

Annexin I from *Bos taurus*, *Homo sapiens*, *Mus musculus*, *Sus scrofa*, *Rattus norvegicus*, *Oryctolagus cuniculus*, *Cavia cutleri*, *Gallus gallus*, and *Columba livia*, respectively;
Annexin III from *Homo sapiens* and *Rattus norvegicus*, respectively;
Annexin V from *Homo sapiens*, *Mus musculus*, *Rattus norvegicus*, and *Gallus gallus*, respectively;
Uteroglobulin from *Oryctolagus cuniculus* and *Lepus capensis*, respectively; and
Phospholipase A2 inhibitor from *Trimeresurus flavoviridis*,
as well as inhibitors homologous therewith isolated from other sources than those explicitly mentioned.

Ribonuclease inhibitors

Barstar from *Bacillus amyloliquefaciens*;
Ribonuclease inhibitor from *Homo sapiens*, *Sus scrofa*, and *Rattus norvegicus*, respectively, as well as inhibitors homologous therewith isolated from other sources than those explicitly mentioned.

RNA polymerase inhibitors

Bacterial RNA polymerase inhibitors from Bacteriophages T3 and T7 as well as inhibitors homologous therewith isolated from other sources.

DNA-entry nuclease inhibitors

Competence protein J from *Bacillus subtilis* as well as inhibitors homologous therewith isolated from other sources.

Beta-lactamase inhibitors

Beta-lactamase inhibitor from *Streptomyces clavuligerus* as well as inhibitors homologous therewith isolated from other sources.

To this list of enzyme inhibitors can be added functionally related inhibitors such as the inhibitor of calcium transport seminalplasmin from *Bos taurus* and *Mus musculus*, respectively, as well as inhibitors homologous therewith isolated from other sources.

To the above list of suitable scaffold molecules to be used in the invention should be added other agents having an effect on the activity of enzymes. One interesting candidate is thioredoxin.

All of the above-listed scaffold molecules can be substituted with an effective part of the complete molecule.

Properties of the scaffold molecule

Since the peptide library presumably can reach every compartment of a cell, it is beneficial if the scaffold enzyme inhibitor protein is not too large, and that it is stable towards e.g. proteolytic attack and insensitive to the reducing environment inside eukaryotic cells. Hence, its function should preferably not be dependent on the formation of disulphide bridges, since these are not formed in the cytosol or nucleus of such cells. In addition the scaffold protein should contain one or more exposed loops in which peptides can be inserted without markedly changing the structure or stability of the protein. On this basis the chymotrypsin inhibitor 2A (cf. the discussion below) is a suitable scaffold but other scaffolds such as the examples mentioned below could also be used.

The members of the peptide library introduced into target cells can effect change of the target cells' phenotype either by direct or indirect (transdominant) interaction. However, the method of the invention renders possible the identification of not only the peptide or ribonucleic acid sequence which is responsible for the direct or indirect effect on phenotype, but also allows for the purification, isolation and identification of the target molecule with which the peptide interacts.

In the present specification the barley-derived chymotrypsin inhibitor 2A has been used by way of example, mainly because this protein is very well-characterized and extremely stable. It will be understood, however, that the use of this enzyme inhibitor in the Examples in no way should limit the scope of the present invention.

Barley chymotrypsin inhibitor 2A belongs to a large family of homologous protease inhibitors mainly found in plants. This family includes barley chymotrypsin inhibitors 1A, 1B, 2A and 2B, potato inhibitor I, wound-induced tomato inhibitor I, ethylene-responsive tomato inhibitor, wild tomato fruit inhibitor I, a subtilisin inhibitor from broad bean, adzuki bean subtilisin inhibitor, pumpkin trypsin/Hageman factor inhibitor, bitter gourd inhibitor, protoplast-specific trypsin inhibitor from Nicotiana sylvestris, tobacco subtilisin inhibitor, amaranth trypsin/subtilisin inhibitor, and beach canavalia subtilisin inhibitor. The only member of this inhibitor family that is of non-plant origin is the leech elastase/cathepsin G inhibitor eglin C.

CI-2A was originally purified from the endosperm of the high-lysine barley Hiproly and shown to be a tight binding inhibitor of the microbial subtilisins Carlsberg and Novo as well as of chymotrypsin. The N-terminal amino acid residue in CI-2A has a blocked amino group as direct amino acid sequencing was unsuccessful. However, most of the amino acid sequence of CI-2A has been determined at the protein level. In addition, it was shown that CI-2A purified from barley is N-terminally processed either during synthesis and storage in the endosperm or in the process of purification. The absence of the 17 N-terminal amino acid residues does not influence the complex formation of CI-2A with subtilisins. Combining the results of amino acid sequencing and cDNA sequencing it has been deduced that CI-2A consists of 83 amino acid residues in a single polypeptide chain containing no disulfide bonds. The blocked N-terminal amino acid residue is serine. The reactive site in CI-2A has been determined to be the Met59-Glu60 peptide bond and the residues in the region Ile56-Arg62 have been demonstrated to be involved in the intermolecular contacts between inhibitor and protease.

The three-dimensional structures of uncomplexed CI-2A as well as of CI-2A complexed with subtilisin Novo are known from X-ray crystallography. The three-dimensional structure of CI-2A has also been determined using NMR spectroscopy, revealing that the reactive loop of CI-2A is dynamic. CI-2A consists of a single α-helix docking against four β-strands. The surface loop stretches across the free side of the sheet and is composed of eight residues: Gly54-Tyr61. In contrast to most enzyme inhibitors, CI-2A lacks disulfide bonds as well as glycosylation sites. In the structures determined, only the 64 C-terminal amino acid residues are defined (L20-G83); this truncated version retains the functionality of the native protein. Comparing the complexed form with the two uncomplexed forms of CI-2A reveals few differences. The most notable difference is that the reactive site loop seems to have a less ordered structure in the uncomplexed forms than in the complexed form. The three-dimensional structure of CI-2A in complex with subtilisin Novo has also been compared to the three-dimensional structure of eglin C in complex with subtilisin Carlsberg. The two homologous inhibitors have highly similar secondary and tertiary structures.

Recombinant variants of N-terminally truncated CI-2A (CI-2A(L20-G83)) and CI-2A with an N-terminal Asp-Pro extension have been widely used to study the folding and stability of CI-2A. The structure of CI-2A in complex with subtilisin Novo has revealed that the number of intermolecular contacts between inhibitor and protease in the P4-P1 region (Ile56-Met59) of the inhibitor are much larger than in the P1'-P3' region (Glu60-Arg62).

Further aspects of the invention

Having identified a modulator according to the above-detailed methods of the invention it is normally of interest to provide large quantities thereof for the purposes of further research

and development, including possible identification of the target molecule with which the modulator physically interacts.

Therefore, the invention also pertains to a method for the preparation of a replicable expression vector, the method comprising the steps of identifying a modulator by the methods of the invention, and subsequently

- isolating or synthesizing a nucleic acid sequence which encodes the modulator, and
- engineering a replicable expression vector comprising an operon which comprises, in the 5'-3' direction and in operable linkage, 1) a promoter for driving expression of the nucleic acid sequence, 2) optionally a nucleotide sequence encoding a leader peptide, 3) the nucleic acid sequence, and 4) a termination signal.

Such methods are widely known in the field of genetic engineering and molecular biology. The skilled person will find suitable guidance in e.g. Sambrook J, Fritsch EF, Maniatis T. 1989. "Molecular cloning: A laboratory manual", 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

In this part of the invention it is preferred that the promoter is inducible; constitutive promoters are not excluded though. Depending on the choice of host cell to carry and express the vector, the promoter is selected from the group consisting of a bacterial promoter, a fungal promoter such as a yeast promoter, and a mammalian promoter. The vector can be in the form of a plasmid, a phage, a cosmid, a minichromosome, or a virus, again depending on choice of host cell and other considerations applying for the specific case. However, it is most preferred that the expression vector is capable of being integrated into the genome of a suitable host cell, since the expression therein will then be more stable over time than is the case with non-chromosomal transformation of the host cell.

Well-known vector systems are based on bacterial plasmid pBR322, λ-phage, and yeast plasmid YRp7, but other suitable and feasible choices are known to the skilled person.

After having provided a suitable expression vector as outlined above, it is preferred that the modulator is produced by transforming a suitable host cell with an expression vector prepared as described above. Such a host cell can be bacterial (e.g. *E. coli*, *Bacillus subtilis*, or any other suitable bacterial host cell), a fungus (e.g. a yeast cell such as *Saccharomyces cerevisiae* or Piccia pastoris) or a plant, insect or mammalian cell (which can be any of the above-discussed cells or cell-types suitable for use as the "substantially identical cells" in the method of the invention).

After having provided a producer cell line as described above it is now possible to produce the modulator of the invention by growing the transformed cell prepared as described above in a culture medium under conditions which facilitates expression by the cell of the randomly modified nucleotide sequence, and subsequently harvesting the expression product from the cell and/or the culture medium. Alternatively, the modulator can be produced by synthesizing the modulator by means of chemical synthesis on the basis of the sequence determined in step (e). In the case that the modulator is a peptide, the well-known techniques of solid- or liquid-phase peptide synthesis can be employed and also if the modulator is a ribonucleic acid, methods for synthetic production thereof are readily available.

An important part of the invention pertains to the isolation/identification of the target biomolecules which are engaging the modulator, the identification, isolation and production of which is described above. Hence an important part of the invention pertains to a method for isolating a target biomolecule, the method comprising providing a modulator according to

the methods described herein and subsequently using the modulator as an affinity ligand in an affinity purification step so as to isolate the target biomolecule from a suitable sample. The affinity purification step can e.g. be an affinity chromatographic step, an affinity mass spectrometry step, or a co-immunoprecipitation step. However, any suitable method for affinity-based isolation/purification can be employed.

Alternatively, the modulator can be used as a probe against a cDNA library derived from the substantially identical cells or as bait in a two- or three-hybrid system (bacterial, fungal or mammalian).

It is preferred that the target biomolecule is a peptide or a nucleic acid, since this also allows of sequence determination thereof.

The potential use of such a target biomolecule is to employ it in a drug-development program. For this purpose it is often useful to resolve the 3-dimensional structure of the target biomolecule (by means of methods available, e.g. X-ray diffraction studies, NMR analysis, etc).

Having isolated a target biomolecule as described above, the invention allows for the rational selection of a chemical compound to be used as a lead compound in drug development, the method comprising the steps of

- assaying a library of chemical compounds for interaction with a target biomolecule which has been *de novo* isolated according to the methods of the, and
- selecting compounds which interact significantly with the target biomolecule.

This approach can be refined by initially identifying the members of the library by methods of structure-based or non-structure based computer drug-modelling. Suitable non-structure based methods are disclosed in US 5,307,287 and US 5,025,388; a method known as CoMFA). An alternative is HASL (Hypothetical Active Site Lattice; Hypothesis Software). Both these methods are based on 3D-QSAR. A feasible structure-based approach is e.g. disclosed in WO 95/06293.

Finally, a very important part of the present invention pertains to a method for the preparation of a medicinal product, the method comprising the steps of

a) selecting a lead compound by the methods of the invention described above,

b) performing pre-clinical tests with the lead compound in order to assess the suitability thereof as a medicinal product,

c) entering, if the lead compound is deemed suitable in step (b), clinical trials using the lead compound in order to obtain market authorization for a medicinal product including the lead compound as a pharmaceutically active substance, and

d) upon grant of a market authorization, admixing the lead compound with a pharmaceutically acceptable carrier, excipient or diluent and marketing the thus obtained medicinal product.

The above-outlined method should of course take into consideration all necessary requirements in order to meet GCP and GMP standards. At any rate, this method is completely depending on the initial provision of the modulator identified according to the invention.

LEGEND TO THE FIGURE

Fig. 1: Functional expression of CI-2A in NIH3T3 cells. Subtilisin was incubated with increasing amounts of the indicated cell extracts and subsequently assayed for proteolytic activity. Each reaction with a given extract concentration was determined in triplicate and the shown velocities are based on the mean values.

PREAMBLE TO EXAMPLES

In the following, the present invention is illustrated by way of example wherein CI-2A is used as starting point for the Cell-Screen technique adapted according to the invention so as to allow intracellular expression of a scaffold protein inhibitor which is randomly modified in the active site. This example is non-limiting, in the sense that other suitable protein inhibitors of enzymes could be used instead of CI-2A. The skilled person can readily perform the necessary substitutions and modifications necessary in order to apply the principles exemplified below using other protein inhibitors of enzymes as starting point.

EXAMPLE 1

Expression of CI-2A truncate in NIH-3T3 cells

Construction of CMVbipep/CI-2A

The CMVbipep/CI-2A construct was created by insertion of a CI-2A derived cDNA fragment that encodes amino acids 21-84 from the wild type CI-2A sequence (cf. the numbering in SEQ ID NO: 1) and

an N-terminally located methionine residue into the *Bam*HI and *Xho*I sites within the CMVbipep retroviral vector (a standard retroviral vector described in WO 96/38553, cf. Fig. 1 therein). The 5117 base pair *Bam*HI/*Xho*I fragment of CMVbipep was gel purified and extracted using a QIAquick gel extraction kit (Qiagen) according to the manufacturer's instructions. The fragment of the CI-2A gene was amplified from a plasmid containing the entire CI-2A encoding DNA sequence by use of the primers

5'-cgggatccatgaaGAcagagtggccagaG and
5'-CGCTCGAGTCAGCCGACCCTGGGGACCT

in a standard polymerase chain reaction (PCR). The amplified DNA fragment was purified using the QUAquick PCR purification kit (Qiagen) according to the manufacturer's protocol and subsequently digested with *Bam*HI and *Xho*I and purified as described above. An approximately 10-fold molar excess of the CI-2A fragment was incubated together with the CMVbipep derived fragment under standard ligation conditions (New England biolabs). Positive clones were identified and confirmed by DNA sequencing.

<u>Transduction of NIH-3T3 cells</u>

CMVbipep and CMVbipepCI-2A were transfected into the BOSC packaging cell line using a $CaPO_4$ co-precipitation method. BOSC packaging cells (also known as BOSC23 cells, cf. WO 94/19478) were diluted to 5×10^5 cells/cm^2 and incubated at standard cell culturing conditions in Dulbecco's modified eagle media (DMEM, GibcoBRL) including supplements (10% fetal calf serum, L-glutamine, 50 mg/ml penicillin/streptomycin (complete DMEM, GibcoBRL). To prepare the BOSC cells for transfection they were washed once in complete DMEM 2 hours before the transfection. The $CaPO_4$ co-precipitated mixtures were prepared by diluting 10 mg CMVbipep/CI-2A or CMVbipep with 10 mg salmon sperm DNA in 450 ml ddH_2O and adding 50 ml 2.5 M $CaCl_2$. These solutions were

slowly added to 500 ml 2 × HEPES-buffered saline pH 7.05 (280 mM NaCl, 1.5 mM Na_2HPO_4, 50 mM HEPES/NaOH pH 7.05) while gently shaking followed by two to five minutes of incubation at 25°C before adding the precipitate to the prepared BOSC cells. After 24 hours of incubation, the cells were washed twice in PBS (137 mM NaCl, 2.7 mM KCl, 8.3 mM Na_2HPO_4 1.4 mM KH_2PO_4) and further cultivated in 10 ml DMEM including supplements for another 24 hours. The media from the transfected BOSC cells were collected and diluted from 10 to 10^5 fold in complete DMEM including 6 mg/ml Polybene. NIH3T3 cells that had been diluted into 10^4 cells/cm^2 and incubated at standard cell culture conditions for 24 hours in complete DMEM were incubated with the virus-containing media for 24 hours at standard conditions. After this incubation period the transduced cell were washed twice in PBS and incubated in DMEM including supplements containing 0.6 mg/ml Geniticin (GibcoBRL).

Preparation of cell extracts

The CMVbipep/CI-2A or CMVbipep transduced NIH3T3 cell were each harvested from two confluent T175 cell culturing bottles by incubation with 5 ml 0.5 × Trypsin-EDTA solution (GibcoBRL)/plate for 10 min at 37°C. The recovered cells were diluted 1:1 in complete DMEM and subsequently collected by centrifugation, washed twice in 10ml DMEM and twice in 1 ml PBS. Finally, the cells were suspended 100 ml PBS. The cells were rendered permeable by three cycles of freezing in liquid nitrogen and thawing by incubation at 37°C and subsequently centrifuged at 20000 × g for 15 minutes cell to remove cell debris. To inactivate endogenous protease activity the extracts were incubated at 65°C for 15 minutes and re-centrifuged.

Subtilisin assay

The indicated cell extracts were diluted to the indicated amounts in 10 ml PBS and added to 25 ml 0.1 M Tris/HCl pH 8.6 containing 5×10^{-8} M Subtilisin Carlsberg (Sigma). After incubation at 25°C for 30 minutes, 25 ml 0.1 M Tris/HCl pH 8.6 containing 5 mM N-succinyl-Ala-Ala-Pro-Phe-p-Nitroanilide (Sigma) was added and the substrate conversion was followed by measuring the absorbance at 405 nm every two minutes until the reaction had reached exhaustion.

Results/discussion

A potential scaffold for intracellular presentation of a peptide library expressed by virtue of a retroviral vector should not have any significant effects on neither the retroviral replication cycle nor the viability of mammalian cells. Furthermore, it must retain a stable structure. In order to test whether CI-2A fulfill these requirements CMVbipep/CI-2A was constructed consisting of the cDNA that encodes the CI-2A amino acids 21-84 inserted into the CMVbipep retroviral vector. This CI-2A expression construct was in parallel with CMVbipep transiently transfected into BOSC packaging cells and the produced viruses were transduced into NIH-3T3 cells.

The titer of both CMVbipep and CMVbipep/CI-2A, determined as geneticin resistant colonies, were in both cases found to be 10^5 infectious virions/ml media. This demonstrated that the presence of the CI-2A gene within the viral genome did not affect the viral replication cycle to any detectable extent.

To detect whether any expression of CI-2A occurs in the transduced cells the potential of cell extracts obtained from the NIH3T3 cells transduced with either CMVbipep or CMVbipep/CI-2A to inhibit the proteolytic activity of subtilisin was determined. This was done by incubation of various amounts of the cell extracts with subtilisin and subsequently detecting the

residual protease activity using a chromogenic substrate (cf. Fig. 1). A complete block of subtilisin activity was observed when the reaction was performed in the presence of between 10 and 0.2 ml of the extract derived from the CMVbipep/CI-2A transduced NIH3T3 cells whereas approximately 50% of the subtilisin activity was maintained when 0.1 ml extract was present. If the amount of CMVbipep/CI-2A extract was further decreased, the subtilisin activity was not detectably reduced. The specificity of the CI-2A mediated inhibitory effect was demonstrated by the observation that even as much as 10 ml of extract from CMVbipep transduced control NIH3T3 cells had no effect on the subtilisin activity.

The data presented demonstrates that CI-2A does not affect the viral packaging and integration processes and that it can be expressed in detectable amounts from the CMVbipep vector. The ability of CI-2A to inhibit the proteolytic activity of subtilisin furthermore suggests that CI-2A adopts a functional structure within mammalian cells. This is of great importance for the use of CI-2A as a scaffold since it then becomes possible to utilise the impressive amount of structural information available on CI-2A to ensure that a peptide library is properly presented. The possibility of expanding the CMVbipep transduced NIH3T3 cell indicates that the presence of CI-2A does not affect to cell viability, further supporting that CI-2A is a suitable scaffold for intracellular presentation of peptides.

EXAMPLE 2

Chymotrypsin inhibitor 2A (CI-2A) as scaffold in the CellScreen technology

As demonstrated in Example 1, CI-2A can be expressed in a functional form in mammalian cells. The provision of a functional

system for displaying randomized peptide sequence using CI-2A as a scaffold is thus relatively uncomplicated to envisage.

It has been shown that insertions of 7, 9, 11 and 13 residues between the Met40 and Glu41 (corresponding to the Met59 and Glu60 in the native molecule) in CI-2A have a minimal effect on the stability and folding rates of the protein. The loop therefore seems to be suitable for insertions of residues for the purposes of the present invention. Moreover, the truncated form of CI-2A has been found to fold by interactions of key residues in the - and C-terminal domains of the protein irrespective of whether the loop is cleaved.

CI-2A is thus an extremely stable protein that has the advantage of being small, having no disulfide bridges, no glycosylation sites and best of all - a loop of eight amino acids that can be manipulated without any major stability loss. To this should be added that CI-2A is of plant origin, and therefore its use in prokaryotic and mammalian systems will not be obscured by the presence of highly homologous proteins.

Especially the lack of disulphide bridges qualifies CI-2A as a scaffold for displaying diversity in the nucleus and/or cytosolic compartment of the cell. As a scaffold, the 64 residue version of CI-2A could be used, since the N-terminal 19 amino acid residues do not have any known function. Peptide libraries could then be inserted into the active-site loop or several amino acid residues in the loop could be substituted with a peptide library. A peptide library inserted in the active-site loop of a protease inhibitor would very likely contain protease inhibitors of several new specificities but, as detailed above, also sequences reactive with other biomolecules *in vivo*.

By fusing the C-terminal of the truncated CI-2A to the surface protein of the bacteriophage pIII, it is contemplated to use

CI-2A scaffolds for selection by phage display, which would make it possible to enrich the library for binders to intracellular components before expressing the library intracellularly in mammalian cells. In order to direct the CI-2A scaffold to different compartments of the cell retroviral vectors holding different leader sequences will be constructed, cf. the above discussion.

To be able to isolate the CI-2A scaffold - and thereby the target molecule to which it binds - from selected cells, a peptide tag will be engineered to the N-terminus of the truncated CI-2A. We are currently considering the following tags: His-tag, Strep-tag or FLAG-tag. However, for future use we are planning to use antibodies with specificity to CI-2A or to a peptide sequence of CI-2A. Peptide libraries of different sizes (six- to nine-mers) would either be inserted between Met40 and Glu41 or amino acid residues in the loop would be substituted with a peptide library.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT:
(A) NAME: M&E Biotech A/S
(B) STREET: Kogle Allé 6
(C) CITY: Hørsholm
(E) COUNTRY: Denmark
(F) POSTAL CODE (ZIP): DK-2970
(G) TELEPHONE: +45 45 16 25 25
(H) TELEFAX: +45 45 16 25 00

(ii) TITLE OF INVENTION: Novel method for the identification of ligand and target biomolecules

(iii) NUMBER OF SEQUENCES: 2

(iv) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 451 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: cDNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Hordeum vulgare
(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION:85..336

(ix) FEATURE:
(A) NAME/KEY: mat_peptide
(B) LOCATION:88..336

(ix) FEATURE:
(A) NAME/KEY: Duplication
(B) LOCATION: 249
(D) OTHER INFORMATION: G is G or T

(ix) FEATURE:
(A) NAME/KEY: Duplication
(B) LOCATION: 252

(D) OTHER INFORMATION: A is A or C

(ix) FEATURE:
(A) NAME/KEY: Duplication
(B) LOCATION: 279
(D) OTHER INFORMATION: C is C or T

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

```
CATTAAACTG ATGACATGAC AGTTCAAGAT CTCACAGTCA CATCGGCGAT CTAATCAGTC       60

TTACAGGAAG CGAGCGTAAC AATG ATG AGT TCA GTG GAG AAG AAG CCG GAG        111
                           Met Ser Ser Val Glu Lys Lys Pro Glu
                            1   1               5

GGA GTG AAC ACC GGT GCT GGT GAC CGT CAC AAC CTG AAG ACA GAG TGG        159
Gly Val Asn Thr Gly Ala Gly Asp Arg His Asn Leu Lys Thr Glu Trp
     10                  15                  20

CCA GAG TTG GTG GGG AAA TCG GTG GAG GAG GCC AAG AAG GTG ATT CTG        207
Pro Glu Leu Val Gly Lys Ser Val Glu Glu Ala Lys Lys Val Ile Leu
 25                  30                  35                  40

CAG GAC AAG CCA GAG GCG CAA ATC ATA GTT CTG CCG GTG GGG ACA ATT        255
Gln Asp Lys Pro Glu Ala Gln Ile Ile Val Leu Pro Val Gly Thr Ile
                 45                  50                  55

GTG ACC ATG GAA TAT CGG ATC GAC CGC GTC CGC CTC TTT GTC GAT AAA        303
Val Thr Met Glu Tyr Arg Ile Asp Arg Val Arg Leu Phe Val Asp Lys
             60                  65                  70
CTC GAC AAC ATT GCC CAG GTC CCC AGG GTC GGC TAGCAAGCTT GAGAGCTAGC      356
Leu Asp Asn Ile Ala Gln Val Pro Arg Val Gly
         75                  80

CTGCTGCTGG CGTGTATGTA TTGCAGCTTC ACCATCTCTT CTTGGCTATA GCAAGATTGA      416

GATTTATAAA TCATATACAA TAAGAGTTGC TGCGG                                 451
```

(2) INFORMATION FOR SEQ ID NO: 2:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 84 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

```
Met Ser Ser Val Glu Lys Lys Pro Glu Gly Val Asn Thr Gly Ala Gly
 -1   1               5                  10                  15

Asp Arg His Asn Leu Lys Thr Glu Trp Pro Glu Leu Val Gly Lys Ser
                 20                  25                  30
```

```
Val Glu Glu Ala Lys Lys Val Ile Leu Gln Asp Lys Pro Glu Ala Gln
            35                  40                  45

Ile Ile Val Leu Pro Val Gly Thr Ile Val Thr Met Glu Tyr Arg Ile
        50                  55                  60

Asp Arg Val Arg Leu Phe Val Asp Lys Leu Asp Asn Ile Ala Gln Val
    65                  70                  75

Pro Arg Val Gly
80
```

CLAIMS

1. A method for identifying an in vivo active modulator of activity of a target enzyme, the method comprising the steps of

(a) preparing a pool of expression vectors, each vector of said pool containing at least one member from a library of randomly modified nucleotide sequences derived from a parent nucleotide sequence encoding a parent peptide or parent ribonucleic acid which modulates the target enzyme activity,

(b) transforming a population of substantially identical cells with said vectors of said pool so as to obtain transformed cells, said substantially identical cells being ones which harbour the target enzyme,

(c) culturing said transformed cells under conditions facilitating expression of said randomly modified nucleotide sequences,

(d) examining said transformed cells and isolating transformed cell(s) wherein the activity of the target enzyme is modulated, and

(e) determining said randomly modified nucleotide sequence of said vector present in cell(s) isolated in step (d) and/or determining the amino acid sequence or the ribonucleic acid sequence of the expression product encoded by said randomly modified nucleotide sequence.

2. The method according to claim 1, wherein the randomly modified nucleotide sequences consist of 1) an invariable part of the parent nucleotide sequence, and 2) random nucleotides.

3. The method according to claim 2, wherein the invariable part of the parent nucleotide sequence encodes a scaffold portion of the parent peptide or of the parent ribonucleic acid which serves to stabilize said polypeptide fragment or ribonucleic acid fragment.

4. A method for identifying a modulator in the form of a biologically active polypeptide fragment or ribonucleic acid fragment which is capable of detectably modulating, in vivo, a phenotypic trait in a cell, the method comprising the steps of

(a) preparing a pool of expression vectors, each vector of said pool containing at least one member from a library of randomly modified nucleotide sequences derived from a parent nucleotide sequence encoding a parent peptide or parent ribonucleic acid which in vivo modulates activity of a known enzyme, wherein the randomly modified nucleotide sequences comprise
- an invariable part encoding a scaffold portion of the parent peptide or of the parent ribonucleic acid, said scaffold portion serving to stabilize said polypeptide fragment or ribonucleic acid fragment, and
- random nucleotides,

(b) transforming a population of substantially identical cells with said vectors of said pool so as to obtain transformed cells,

(c) culturing said transformed cells under conditions facilitating expression of said randomly modified nucleotide sequences,

(d) examining said transformed cells and isolating transformed cell(s) wherein the preselected phenotypic trait is modu-

lated by the presence of the expressed randomly modified nucleotide sequence, and

(e) determining said randomly modified nucleotide sequence of said vector present in cell(s) isolated in step (d) and/or determining the amino acid sequence or the ribonucleic acid sequence of the expression product encoded by said randomly modified nucleotide sequence.

5. The method according to any of the preceding claims wherein the substantially identical cells are prokaryotic cells.

6. The method according to claim 5, wherein the prokaryotic cells are bacteria selected from the group consisting of Bacillus spp., Clostridium spp., Corynebacterium spp., Staphylococcus spp., Streptococcus spp., Escherichia coli, Serratia marcescens, Klebsiella spp., Proteus spp., Citrobacter spp., Salmonella spp., Shigella spp., Pseudomonas spp., Acinetobacter spp., Aeromonas spp., Plesiomonas spp., Yersinia spp., Francisella tularensis, Vibrio spp., Campylobacter spp., Helicobacter pylori, Haemophilus spp., Bordetella spp., Brucella spp., Neisseria spp., Treponema pallidum, Leptospira interrogans; Borrelia spp., Legionella pneumophila, Listeria monocytogenes, Mycobacterium spp., Treponema pallidum, Chlamydia trachomatis, Actinomyces spp., Rickettsia spp., Mycoplasma spp., Lactococcus spp., and Lactobacillus spp.

7. The method according to claim 6, wherein
the Clostridium spp. is selected from the group consisting of C. botulinum, C. difficile, C. perfringens, and C. tetani
the Corynebacterium spp. is selected from the group consisting of C. diphtheriae, and C. pyogenes;
the Bacillus spp. is selected from the group consisting of B. anthracis, B. subtilis and B. cereus;

the Staphylococcus spp. is selected from the group consisting of S. aureus and S. albicans;
the Streptococcus spp. is selected from the group consisting of S. pneumoniae, S. pyogenes, and S. agalactiae;
the Klebsiella spp. is K. pneumoniae;
the Proteus spp. is P. mirabilis;
the Citrobacter spp. is Citrobacter freundii;the Salmonella spp. is selected from the group consisting of S. typhi, S. typhimurium, S. shottmülleri and S. paratyphi;
the Shigella spp. is selected from the group consisting of S. dystenteriae, S. flexneri, S. boydii, and S. sonnei;
the Pseudomonas spp. is selected from the group consisting of P. aeruginosa, P. pseudomallei, and P. mallei;
the Yersinia spp. is selected from the group consisting of Y. pestis, Y. enterocolitica, and Y. pseudotuberculosis;
the Vibrio spp. is selected from the group consisting of V. cholerae and V. parahaemolyticus;
the Campylobacter spp. is selected from the group consisting of C. jejuni and C. coli;
the Haemophilus spp. is selected from the group consisting of H. influenzea, H. parainfluenzae, and H. aegyptius;
the Bordetella spp. is selected from the group consisting of B. pertussis, B. parapertussis, and B. bronchiseptica;
the Neisseria spp. is selected from the group consisting of N. gonnorhoeae and N. meningitidis;
the Borrelia spp. is selected from the group consisting of B. burgdorferi sensu stricto, B. garinii, B. afzelii, and B. recurrentis;
the Mycobacterium spp. is selected from the group consisting of M. tuberculosis, M. bovis, M. africanum, M. kansasii, and M. leprae; and
the Mycoplasma spp. is M. pneumoniae.

8. The method according to any one of claims 1-4, wherein the substantially identical cells are eukaryotic cells.

9. The method according to claim 8, wherein the eukaryotic cells are selected from the group consisting of fungal cells, protozoan cells, animal cells, and plant cells.

10. The method according to claim 9, wherein the fungal cells are selected from the group consisting of Epidermophyton spp., Trichophyton spp., Microsporum spp., Candida albicans, Philophora spp., Coccidioides immitis, Histoplasma capsulatum, Blastomyces dermatitidis, Paracoccidioides brasiliensis, Cryptococcus neoformans, Aspergillus spp., Saccharomyces cerevisiae, Klyveromyces lactis, and Piccia pastoris.

11. The method according to claim 9, wherein the protozoan cells are selected from the group consisting of Giardia lamblia, Trichomonas vaginalis, Dientamoeba fragilis, Trypanosoma spp., Leishmania spp., Entamoeba histolytica, Naegleria fowleri, Acanthamoeba castellani, Harmanella spp., Isospora belli, Cryptosporidium spp., Sarcocystis spp., Toxoplasma gondii, Plasmodium spp., Babesia spp., and Balantidium coli.

12. The method according to claim 9, wherein the plant cells are derived from a plant selected from the group consisting of Nicotiana tabacum (tobacco plant), Arabidopsis thaliana, Brassica napus, Brassica juncea, Musa sp. (banana plant), rice, and corn.

13. The method according to claim 9, wherein the animal cells are selected from the group consisting of mammalian cells, arthropod cells such as insect cells, avian cells, and piscine cells.

14. A method according to claim 13, wherein the mammalian cells are selected from the group consisting of Chinese hamster ovary (CHO) cells, VERO cells, HeLa cells, W138 cells, BHK cells, COS-7 293 cells, and MDCK cells.

15. The method according to any of the preceding claims, wherein the transformed cells examined in step (d) predominantly carries one single copy of the vector.

16. The method according to claim 15, wherein transformation step (b) is performed under such conditions that the cells transformed are predominantly or at most transformed with one single vector from said pool, or wherein, prior to carrying out step (d), cells being transformed with more than one vector from said pool are substantially excluded from the further steps.

17. The method according to any of the preceding claims, wherein the modulator is a peptide.

18. The method according to any of claims 1-16, wherein the modulator is a nucleic acid fragment such as an RNA fragment.

19. The method according to any of the preceding claims, wherein the modulator is stable towards proteolytic attack and/or is insensitive to a reducing environment.

20. The method according to any one of claims 2-19, wherein the random nucleotides are introduced in part(s) of the parent nucleotide sequence which encode(s) the active site(s) of the parent peptide or parent ribonucleic acid, or the part(s) which encode(s) structure(s) interfering with the active site(s).

21. The method according to any one of claims 2-20, wherein the invariable part of the nucleotide sequence encodes truncated parts of the parent peptide or parent ribonucleic acid sufficient to maintain stability.

22. The method according to any of claims 2-21, wherein the invariable part of the parent nucleotide sequence encodes a peptide which is free from disulphide bridges.

23. The method according to any one of claims 2-21, wherein the invariable part of the parent nucleotide sequence encodes a peptide having disulphide bridges.

24. The method according to any one of claims 2-23, wherein the random nucleotides are introduced in the form of an insertion or a substitution into the parent nucleotide sequence, optionally in combination with deletion(s) in the parent nucleotide sequence.

25. The method according to claim 24, wherein the number of random nucleotides which are introduced is in the range from 3 to about 100.

26. The method according to any one of claims 2-25, wherein the random nucleotides are nucleotide sequences and/or are single random nucleotides introduced at specific sites in the parent nucleotide sequence.

27. The method according to any one of claims 2-26, wherein the random nucleotides are selected from the group consisting of

synthetic, completely random deoxyribonucleotides;

synthetic random DNA sequences, wherein limitation on randomization of some nucleotides is introduced so as to limit the number of available sequences and/or to avoid undesired stop codons and/or to facilitate introduction of post-translational modifications of expressed peptide(s);

synthetic random DNA sequences as in (1) or (2) coupled to a sequence encoding a purification tag; and

CDR encoding nucleotide sequences isolated from a library of immune-competent cells raised against an antigen.

28. The method according to claim 27, wherein the CDR encoding nucleotide sequences encode CDR-3 peptide sequences.

29. The method according to any one of claims 26-28, wherein the random nucleotides are prepared by codon split synthesis where defined DNA codons are synthesized in a random order.

30. The method according to claim 29, wherein the relative amount of synthesized codons ensure that all encoded amino acids will be present with substantially the same frequency in encoded polypeptide fragments.

31. The method according to any one of claims 2-30, wherein the random nucleotides are introduced into the expression vector by the principle of site directed PCR-mediated mutagenesis.

32. The method according to any one of the preceding claims, wherein the modulator in vivo reduces KM of the target enzyme for at least one substrate.

33. The method according to any one of claims 1-31, wherein the modulator in vivo increases KM of the target enzyme for at least one substrate.

34. The method according to any of the preceding claims, wherein the modulator in vivo increases Vmax of the target enzyme for at least one substrate.

35. The method according to any of claims 1-33, wherein the modulator in vivo reduces Vmax of the target enzyme for at least one substrate.

36. The method according to any one of the preceding claims, wherein the parent peptide or parent ribonucleic acid is an inhibitor of activity of the target enzyme.

37. A method according to claim 36, wherein the inhibitor is selected from the group consisting of a BPTI/Kunitz family protease inhibitor, a serpin family protease inhibitor, a Kazal family protease inhibitor, a soybean trypsin inhibitor (Kunitz) family protease inhibitor, a potato inhibitor I family member, a Bowman-Birk family protease inhibitor, a squash inhibitor family member, a wap-type 'Four-disulfide Core' proteinase inhibitor, a hirudin family protease inhibitor, a factor Xa inhibitor, an Ascaris trypsin inhibitor family member, a cystatin family protease inhibitor, a calpain family cysteine protease inhibitor, a tissue inhibitor of metalloproteinases family member, a carboxypeptidase A inhibitor, a metallocarboxypeptidase inhibitor, an angiotensin-converting enzyme inhibitor, a cereal alpha-amylase/trypsin inhibitor family member, an alpha-amylase/trypsin inhibitor homologous to thaumatin, an alpha-amylase/subtilisin inhibitor family member, an inhibitors of insect alpha-amylases, an inhibitor of mammalian alpha-amylases derived from Streptomyces species, a trehalase inhibitor, a polygalacturonase inhibitor, a fucosyltransferase inhibitor, a protein kinase C inhibitor, an cAMP-dependent protein kinase inhibitor, a cyclic nucleotide phosphodiesterase inhibitor, a protein phosphatase inhibitor, a TCD/MRS6 family GDP dissociation inhibitor, an ATPase inhibitor, a phospholipase A2 inhibitory protein, a ribonuclease inhibitor, an RNA polymerase inhibitor, a DNA-entry nuclease inhibitor, and a beta-lactamase inhibitor.

38. The method according to any of the preceding claims wherein the substantially identical cells are mammalian cells and the vector is selected from the group consisting of a retroviral vector, a vaccinia virus vector, an adenoviral vector, an adeno associated virus (AAV) vector, a herpes simplex virus (HSV) vector, an alpha virus vector, and a semliki forest virus vector.

39. The method according to claim 38, wherein the vector is retroviral.

40. The method according to claim 39, wherein the retroviral vector is derived from retrovirus selected from the group consisting of Avian Leukosis-Sarcoma Virus (ASLV), Mammalian type C, Mammalian type B, and Lentivirus, and optionally modified with heterologous cis-acting elements.

41. The method according to claim 39 or 40, wherein the retroviral vector has non-identical ends.

42. The method according to claim 41, wherein the non-identical ends contain non-identical promoters.

43. The method according to any one of claims 39-42, wherein the retroviral vector contains heterologous promoter replacing the viral promoter in the 5'-LTR, such as a CMV promoter, an RSV promoter, an SV-40 promoter, a TK promoter, an MT promoter, or an inducible system such as Tet or Ecdysone.

44. The method according to any one of claims 39-43, wherein step (a) is carried out by

1) transfecting suitable packaging cells with vectors comprising the randomly modified nucleotide sequences and which are integratable in virions produced by said packaging cells,
2) culturing said transfected packaging cells in a culture medium under conditions which facilitates production by the packaging cells of virions containing the randomly modified nucleotide sequences,

3) recovering and optionally concentrating said virions, and

4) transducing said substantially identical cells with the virions.

45. The method according to claim 44, wherein the packaging cells are selected from the group consisting of PE501, Bosc23, Psi2, GP+E86, PhoenixEco, PA317, GP+AM12, DA(ampho), Bing, FLYA13, ProPak, CRIP, PsiAM, Phoenix-Ampho, PG13, and H9 (293GPG).

46. The method according to any of the preceding claims wherein the randomly modified nucleotide sequences are coupled to a nucleotide sequence encoding at least one fusion partner.

47. The method according to claim 46, wherein the fusion partner serves to facilitate expression and/or purification/isolation and/or further stabilization of the expression product.

48. The method according to claim 47, wherein the fusion partner includes a purification tag such as His6 tag, myc tag, BSP biotinylation target sequence, of BirA, flu tag, lacZ, and GST.

49. The method according to claim 46 or 47, wherein the fusion partner is a sorting signal or a targeting sequence.

50. The method according to claim 45, wherein the sorting signal is a signal patch or a signal peptide.

51. The method according to claim 49 or 50, wherein the sorting signal effects export of the expressed peptide out of the cell or into the cell membrane, or, when the substantially identical cells are eukaryotic, into endoplasmic reticulum, into Golgi apparatus, into lysosomes, into secretory vesicles, into mitochondria, into peroxisomes, or into the nucleus.

52. The method any one of claims 49-51, wherein the sorting signal or targeting sequence is selected from the group consisting of

a nuclear localization signal (NLS) such as Pro-Lys-Lys-Lys-Arg-Lys-Val (SV40 large T antigen NLS), Ala-Arg-Arg-Arg-Arg-Pro (human retinoic acid receptor-3 NLS), Glu-Glu-Val-Gln-Arg-Lys-Arg-Gln-Lys-Leu (NFκB p50), Glu-Glu-Lys-Arg-Lys-Arg-Thr-Tyr-Glu (NFκB p65), and Ala-Val-Lys-Arg-Pro-Ala-Ala-Thr-Lys-Lys-Ala-Gly-Gln-Ala-Lys-Lys-Lys-Lys-Leu-Asp (Xenopus nucleoplasmin NLS);

a membrane anchoring sequence such as those derived from CD8, ICAM-2, IL-8, CD4, and LFA-1, and a lipidation sequence such as a myristylation or a palmitoylation sequence;

a lysosomal sorting signal such as a lysosomal degradation sequence, and a lysosomal membrane sequencea mitochondrial localization sequence such as a mitochondrial matrix sequence, a mitochondrial inner membrane sequence, a mitochondrial intermembrane space sequence, and a mitochondrial outer membrane sequence; an endoplasmic reticulum localization sequence such as the sequence from calreticulin (KDEL) and the sequence from adenovirus E3/19K protein (LULSRRSFIDEKKMP);

a peroxisome sequence such as the peroxisome matrix sequence from Luciferase;

a farnesylation sequence such as LNPPDESGPGCMSCKCVLS;

a geranylgeranylation sequence such as LTEPTQPTRNQCCSN;

a destruction sequence such as RTALGDIGN; and

a secretory signal sequence such as the secretory signals from IL-2, growth hormone, preproinsulin, and influenza HA protein.

53. The method according to any one of the preceding claims, which further comprises the step of resolving the 3-dimensional structure of the identified modulator.

54. A method for the preparation of a replicable expression vector, the method comprising the steps of identifying a modula-

tor by the method according to any one of claims 1-53, and subsequently

i) isolating or synthesizing a nucleic acid sequence which encodes the modulator, and

ii) engineering a replicable expression vector comprising an operon which comprises, in the 5'-3' direction and in operable linkage, 1) a promoter for driving expression of the nucleic acid sequence, 2) optionally a nucleotide sequence encoding a leader peptide, 3) the nucleic acid sequence, and 4) a termination signal.

55. The method according to claim 54, wherein the promoter is inducible.

56. The method according to claim 55, wherein the promoter is selected from the group consisting of a bacterial promoter, a fungal promoter such as a yeast promoter, and a mammalian promoter.

57. A method according to any of claims 54-56, wherein the expression vector is a plasmid, a phage, a cosmid, a minichromosome, or a virus.

58. A method according to any of claims 54-57, wherein the expression vector is capable of being integrated into the genome of a suitable host cell.

59. A method according to any of claims 54-58, wherein the nucleotide sequence is incorporated into a vector selected from the group consisting of pBR322, phage-λ and YRp7.

60. A method for the preparation of a transformed cell carrying a nucleic acid sequence encoding a modulator as defined in any one of claims 1-53, the method comprising transforming a suit-

able host cell with an expression vector prepared according to any one of claims 54-59.

61. A method for providing a modulator as defined in any one of claims 1-53, the method comprising

I) growing a transformed cell prepared according to the method of claim 60 in a culture medium under conditions which facilitates expression by the cell of the randomly modified nucleotide sequence, and

II) subsequently harvesting the expression product from the cell and/or the culture medium, or

Ia) identifying the modulator according to the method of any one of claims 1-53, and

Ib) subsequently synthesizing the modulator by means of chemical synthesis on the basis of the sequence determined in step (e).

62. A method for isolating a target biomolecule, the method comprising providing a modulator according to the method of claim 61 and subsequently using the modulator as an affinity ligand in an affinity purification step so as to isolate the target biomolecule from a suitable sample.

63. The method according to claim 62, wherein the affinity purification step is an affinity chromatographic step, an affinity mass spectrometry step, or a co-immunoprecipitation step.

64. A method for isolating a target biomolecule, the method comprising providing a peptide modulator according to the method of claim 61 and subsequently using the modulator as a probe against a cDNA library derived from the substantially identical cells or using the modulator as bait in a two- or three-hybrid system.

65. The method according to any of claims 61-63, wherein the target biomolecule is a peptide or a nucleic acid.

66. The method according to claim 64 or 65 further comprising the step of determining the amino acid sequence of the peptide or determining the nucleotide sequence of the nucleic acid.

67. The method according to claim 66, further comprising the step of resolving the 3-dimensional structure of the target biomolecule.

68. A method for selecting a chemical compound as a lead compound in drug development, the method comprising the steps of

- assaying a library of chemical compounds for interaction with a target biomolecule which has been de novo isolated according to the method of any one of claims 62-67, and
- selecting compounds which interact significantly with the target biomolecule.

69. The method according to claim 68, wherein the library of chemical compounds has been provided by chemical synthesis upon initial identification of the members of the library by structure-based or non-structure based computer drug-modelling.

70. A method for the preparation of a medicinal product, the method comprising the steps of

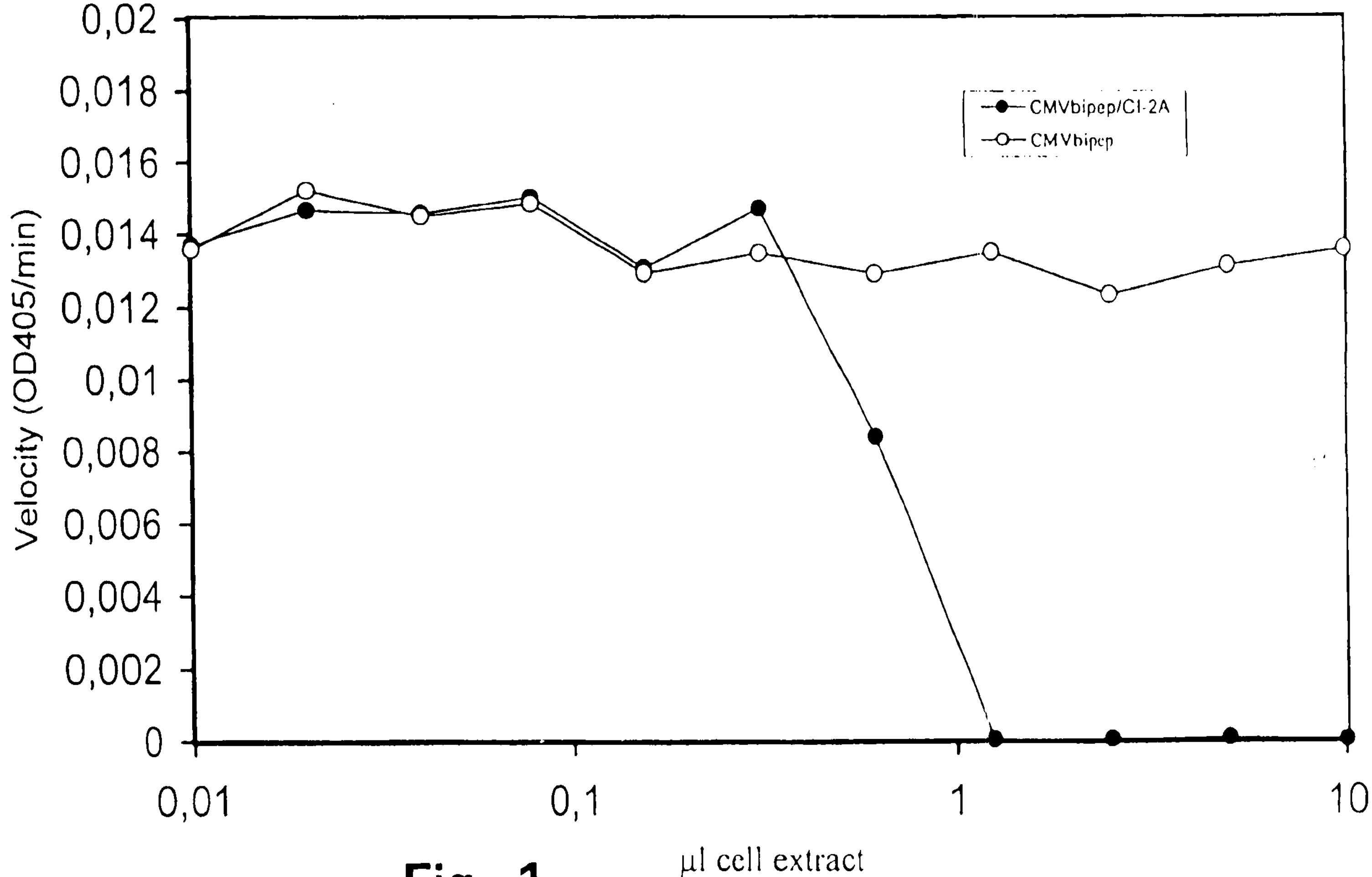
A) selecting a lead compound by the method according to claim 68 or 69,

B) performing pre-clinical tests with the lead compound in order to assess the suitability thereof as a medicinal product,

C) entering, if the lead compound is deemed suitable in step (B), clinical trials using the lead compound in order to

obtain market authorization for a medicinal product including the lead compound as a pharmaceutically active substance, and

D) upon grant of a market authorization, admixing the lead compound with a pharmaceutically acceptable carrier, excipient or diluent and marketing the thus obtained medicinal product.



CMVbipep/CI-2A
CMVbipep
Velocity (OD405/min)
0,02
0,018
0,016
0,014
0,012
0,01
0,008
0,006
0,004
0,002
0
0,01
0,1
1
10
µl cell extract

Fig. 1